Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/004180

International filing date: 11 February 2005 (11.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/543,940
Filing date: 13 February 2004 (13.02.2004)

Date of receipt at the International Bureau: 11 March 2005 (11.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1291120

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/543,940*
FILING DATE: *February 13, 2004*
RELATED PCT APPLICATION NUMBER: *PCT/US05/04180*

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

Please type a plus sign (+) inside this box → [+]

PTO/SB/16 (8-00)
Approved for use through10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Marc V.	Gorenstein	Needham, MA
Robert S.	Plumb	Milford, MA
Chris L.	Stumpf	Uxbridge, MA

☐ *Additional inventors are being named on th____ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (280 characters max)

APPARATUS AND METHOD FOR IDENTIFYING PEAKS IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY DATA AND FOR FORMING SPECTRA AND CHROMATOGRAMS

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☒ Customer Number **28970** →

OR *Type Customer Number here*

Place Customer Number Bar Code Label here

☒ Firm or Individual Name	David C. Issacson				
Address	SHAW PITTMAN LLP				
Address	1650 Tysons Boulevard				
City	McLean	State	VA	ZIP	22102
Country	USA	Telephone	703-770-7900	Fax	703-770-7901

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Number of Pages* 141 ☐ CD(s), Number

☒ Drawing(s) *Number of Sheets* 15 ☒ Other (specify)

☐ Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

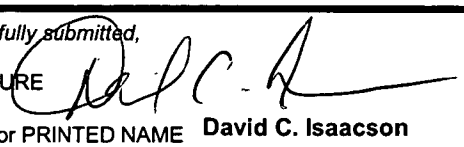
		FILING FEE AMOUNT ($)
☐	A check or money order is enclosed to cover the filing fees	
☒	The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 50-1390	$160.00
☐	Payment by credit card. Form PTO-2038 is attached.	

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE 

Date 2/13/04

TYPED or PRINTED NAME David C. Isaacson

REGISTRATION NO. *(if appropriate)* 38,500

Docket Number: WAA-347-PRO

TELEPHONE 703-770-7900

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

APPARATUS AND METHOD FOR IDENTIFYING PEAKS IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY DATA AND FOR FORMING SPECTRA AND CHROMATOGRAMS

Marc V. Gorenstein, Robert Stephen Plumb, Chris Lee Stumpf

Waters Corporation

Specification

A method of detecting ions and forming mass-to-charge spectra by means of liquid chromatography and mass spectrometry (LC/MS) is disclosed.

Summary of the Invention

The key parameters of ions, which are their mass-to-charge ration (m/z) retention time, and intensity, can be precisely and accurately estimated via the convolution of an LC/MS data matrix using fast, linear, FIR filters, followed by apex detection and location.

These Ion parameters (obtained, preferably from this convolution operation) can be used to reduce complexity in spectra.

Object of Invention

A first object of this invention is the production of a complete accounting of the ions detected by an LC/MS apparatus. The chromatograms produced by an LC/MS apparatus contain noise and co-eluted compounds and partially resolved ions. The detection method described here, based upon linear convolution of the LC/MS data matrix, reduces the effects of noise and resolves partially co-eluted compounds and unresolved ions. The reduction of noise, as provided by this method, results in an increase in the number of ions that are reliably detected. The partial resolution of co-eluted molecules and interfered ions, as provided by this method, further increases the number of ions that are reliably detected. The method results in a tabular list of ions where each ion is described by its mass-to-charge ration, retention time, and intensity. The detection method here obtains values for these parameters that are optimally estimated in the sense that the precision and reproducibility of these parameters is enhanced.

A second object of this invention is to extract from these tables subsets of ions that have desired properties or relationships. For example, it is well known that ions from a common parent molecule will have essentially identical retention time in an LC/MS chromatogram. Extracting those ions that lie within a retention time window about a parent ion will retain related ions and exclude unrelated ions. The result of this extraction method is a spectrum of reduced complexity. These results provide a significant improvement compared to the usual procedure of simply extracting a spectrum (or an average of spectra) from an LC/MS data matrix. Such extracted spectra are contaminated by the ions from the leading or tailing edge of peaks, unrelated to the ions of interest. The ions retained in the windowed spectrum can then be further analyzed by methods known in the prior art. For example these methods can be used to obtain the mass or identity of the common parent molecule.

Window thresholds can also be applied to extract from the list ions of nearly the same mass-to-charge value. This has the effect of producing chromatograms corresponding to a desired mass-to-charge ratio, producing chromatograms of reduced complexity.

Background of the Invention

Mass spectrometers (MS) are widely used to identify and quantify molecular species in a sample. When a sample is introduced into the MS, the molecules are ionized and introduced into a mass analyzer, which measures the mass-to-charge ratio (m/z) and intensity of ions.

A mass spectrometer is limited as to the number of ions it can reliably detect and quantify within a single spectrum. A single complex sample injected into an MS may well produce spectra too complex to interpret or analyze.

A common technique to reduce the complexity of the resulting spectra is to precede MS analyzes with a chromatographic separation. Such a separation can be carried out by a gas chromatography (GC) or liquid chromatography (LC), giving rise to both the GC/MS and LC/MS methods. The applications of interest here are large-molecule, non-volatile analytes that can be dissolved in solvent. Such analytes are best separated by liquid chromatographic techniques. We will henceforth refer only to LC or LC/MS, though the ion detection and analysis method disclosed will apply as well to GC or GC/MS analysis.

In an LC/MS system, the injection of sample occurs at a single moment, the LC subsequently causes the sample to elute over time, and the eluent is continuously introduced into the ionization source of the mass spectrometer. As the separation progresses, the composition of the mass spectrum evolves over time, reflecting the changing composition of the eluent.

At regularly spaced time intervals, a computer-based system samples and records the spectrum seen at that interval in a storage device, such as a hard-disk drive. Generally, it is only after the LC separation is complete, that the acquired spectra are then analyzed. The subject matter of this application relate to methods used to analyze these post-acquisition, stored spectra.

A sample will generally contain more than one molecular species. Biological samples, in particular, may contain 1000's, 10,000's or more molecular species. A molecular species may produce more than one ion. For example, the mass of a peptide depends on the isotopic forms of its nuclei; the electrospray interface can ionize proteins into families of charge states.

Reducing the number of ions simplifies the interpretation of the spectra. For example, peptides or proteins can produce clusters of ions that elute at a common time. Molecules that elute at the same retention can produces clusters that overlap. The interpretation of such clusters is more straightforward if the clusters from the different molecules are separated in time.

The concentration of a species can vary over a wide range. In biological samples, it is often the case that there are more species by number at lower concentrations that at higher concentration. It follows that a large fraction of ions will appear at low

concentration, near the limit of detection of the LC/MS. Again, the problem of detecting low abundance species is simplified if few species are present in the spectrum at any one time, and if the back-ground noise present in the LC/MS chromatogram is as reduced as much as possible.

It is the object of this invention to further reduce the complexity of spectra obtained in an LC/MS separation. This additional reduction in complexity further simplifies the interpretation of spectra and by spectra containing fewer ions and by reducing noise backgrounds, and by partially resolved co-eluted compounds and interfering ions.

The LC/MS method in detail

The method of liquid chromatography followed by on-line mass spectrometry (LC/MS) provides a powerful means to identify and quantify molecular species in a sample. The LC/MS method is able to analyze a wide variety of samples. A given sample can contain a mixture of a few or thousands of molecular species. The molecules themselves can span a wide range of properties and characteristics.

An LC/MS system generally analyzes the content of a single mixture at a time. We may refer to the analysis of a single mixture by a single LC/MS as the analysis of an injection.

Any given sample is generally only one of a set of samples; it is the sample set that represents an experiment from which meaningful results can be obtained. For example, a sample set can contain calibration samples, control samples, and unknown samples that were obtained under a variety of conditions. The desired result from an experiment might be a determination of how the concentration of one analyte of interest has changed between and within the controls and unknowns.

The analysis of a sample set is typically carried out by analyzing each sample in serial order. To measure the reproducibility of the results obtained from a given injection, a typical protocol may require that each sample to be divided and analyzed in replicate; and each sample may be analyzed by different, but nominally equivalent, LC/MS systems.

An LC/MS system generally analyzes the content of a single mixture at a time. Systems that support parallel separations do exist, however the intent of these systems is to provide greater throughput. The results obtained from a single sample are the same, regardless of whether the data were acquired serially or in parallel.

The object of this invention is to describe methods that extract the maximal information content from each single injection, that is, from each analysis of a single sample mixture by a single LC/MS system. As described in the examples given above, the results obtained from each single injection can then be further analyzed by known methods in order to obtain the final results desired from a sample set. The object of this invention is to provide enhanced completeness, accuracy, and reproducibility of the final result by improving completeness, accuracy, and reproducibility of results obtained from a single injection.

The chromatographic separation

An analyst performs an LC/MS analysis by injecting a sample, by manual or automatic means, into the chromatograph. A high pressure stream of chromatographic solvent forces the sample to migrate through the chromatographic column. The column generally contains a packed bed of silica beads to whose surface are bonded molecules that determine the migration velocity of each molecular species. The resulting migration time of a species depends on competitive interactions between that molecule, the solvent, and the beads.

[FIGURE 1. LC system]

As a result of these interactions, a species then migrates through the column and emerges, or elutes, from the column at a characteristic time, conventionally referred to as the molecule's retention time. Once the peak elutes from the column, it can be conveyed to a detector, such as a mass spectrometer.

A retention time is an average time. A molecule that elutes from a column at retention time t actually elutes over a period of time that is centered at time t. The elution profile is termed a chromatographic *peak*. The elution profile of a peak is typically bell-shaped, and has a width. We describe the peak's width by its full width at half height, or half-maximum (fwhm).

The peak width, as measured by fwhm, is independent of the height of the peak and will be, essentially a constant characteristic of a molecule for a given separation method. Ideally, for a given chromatographic method, all molecular species will elute with the same peak width. In practice the peak width will change with retention time; for example, peaks that elute at the end of a separation may have width that are two times wider than those that elute early in the separation. Thus, as measured by fwhm, peak widths may change by up a factor of 2 or 3 or more. The method described here can accommodate the range of peak widths typically encountered in a chromatographic separation.

The chromatographic separation is a continuous process, but a detector that receives the eluent will generally interrogate the eluent at regularly spaced intervals. The rate at which a detector interrogates the eluent is a key system parameter. This rate or interval is conventionally as a sample frequency or period. The chromatographic peak width determines the sample period. The period must be high enough so that the system adequately samples the profile of each peak. Typically, the sample period is set so the detector will make about 5 measurements during the fwhm of a peak.

We will model a peak by a Gaussian profile. For a Gaussian, the fwhm is a factor of approximately 2.35 times the standard deviation σ of the Gaussian, a measure of width preferred by statisticians.

In addition to having a width, a chromatographic peak has a height or area. The height and area are measures of the response of the detector to the molecular species. Generally, the height and area are proportional to the amount or mass of the species injected into the chromatograph. We will use the term intensity to refer to either the height or area and term intensity is intended to refer to a measure of the detector's response to the amount of the species introduced into the LC/MS system.

The details of the chromatographic system and method determine the interval during which peaks can emerge, the retention time of each peak, and the width of each peak.

The mass spectrometric system

In an LC/MS system, the chromatographic eluent is introduced into the mass spectrometer (MS) portion of the apparatus.

[FIGURE 1. MS system]

The functional components that make up an MS apparatus are the de-solvation system, the ionizer, the mass analyzers, the detector, and the data recording and storage systems.

Upon introduction into the MS system, the desolvation system removes the solvent, and the ionizing source ionizes the analyte molecules. The ionized molecules are then conveyed to the mass analyzer portion. The mass-analyzer sorts or filters the molecules by their mass-to-charge ratio. Molecules at each value for m/z are then detected with a detection apparatus. The detector response is proportional to the intensity of ions at each mass-to-charge interval. The intensity as a function of m/z is the mass-to-charge spectrum.

The mass-to-charge spectrum is then recorded by a computer and stored in a storage medium such as hard-disk drive. This spectrum is recorded as an array of values by computer system and is stored for later display and mathematical analysis.

As mentioned above, the elution of molecules from the chromatographic system is a continuous process, and, as in any LC separation, the detector samples the eluent at regularly spaced time interval. In an LC/MS system, the MS collects measures mass spectra at these regularly spaced time intervals. Thus the output of an LC/MS experiment is a series of spectra. Each spectral scan is described by its scan time. Again, the sample period between spectral scans is set to insure that an adequate number of spectra are collected during the elution of each peak.

We will refer to each spectrum as a scan, and each element of the spectrum as a channel. The collected spectra or scans are stored in a storage medium such as hard-disk drive. A typical LC/MS separation results is a series of mass-to-charge spectra stored on a hard-disk drive, or some equivalent storage system.

Mass analyzers measure only the ratio of a molecules molecular weight to its charge. Thus spectra respond only to the mass-to-charge ratio of an analyte. A molecule of molecular weight m and charge z will appear as an ion with a mass-to-charge ratio m/z. We introduce the symbol μ to refer to the mass-to-charge ratio, thus $\mu \equiv m/z$.

These specific functional elements that make up a MS system can vary widely. The methods described here can apply to the wide range of components that can make up an MS system.

We summarize currently known varieties of MS components:

Methods for ionization include electron-impact (EI), electrospray (ES) , and atmospheric chemical ionization (APCI).

Mass analyzers include a quadruple mass analyzer (Q), and time-of-flight (TOF) mass analyzer, and Fourier-transform-based mass spectrometers (FTMS). Mass analyzers can be placed in tandem in a wide variety of conformations, e.g, Q-TOF. Mass analyzers can include on-line collision modification of an already mass-analyzed molecule. E.g. in triple quadrupole, Q1-Q2-Q3, or a Q1-Q2-TOF, based massed analyzers , the second quadruple Q2, can impress accelerating voltages to the ions separated by Q1. These ions, colliding with gas expressly introduced into Q2, are then fragmented and those fragments are further analyzed by Q3 or by the TOF. Typically it is only the ions after Q3 that detected and it their spectra that are recorded. The methods described here can be applied to spectra obtained from all modes of mass-analysis.

After mass-to-charge analysis, the LC/MS apparatus must detect and record the ions. The detection of ions can be performed by a current measuring electrometer, or a single ion counting multi-channel plate (MCP). For specificity, we shall assume that an MCP is employed and the detection of an ion results a specific number of counts.

The post-separation data analysis system

After the chromatographic separation is completed, the analyst uses the post-separation data analysis system (DAS) to analyze the stored spectra. This system, generally implemented by computer software, can accomplish a number of tasks, which include visual display of the spectra or of the chromatograms, or provide for the mathematical analysis of the data. The analyses provided by the DAS include analyses the results obtained from a single injection. The DAS will also allow the results obtained from a set of injections to be viewed and to be further analyzes. Examples of analyses applied to a sample set include the production of calibration curves for analytes of interest, and the detection of novel compounds present in the unknowns, but not in the controls.

The object of this invention relates primarily to that portions of the DAS that analysis the data obtained from an injection; a single LC/MS analysis of a single sample.

The ion signal in an LC/MS experiment

To illustrate the how an ion appears in an LC/MS experiment, we consider a simulation of a sample that produces three ions, designated as ion 1, ion 2, and ion 3 that appear within a limited range of retention time and m/z. We assume that the mass-to-charge ratios of these ions are different, and that these molecular parents of the ions eluted at nearly, but not exactly, the same retention times.

After acquisition by the LC/MS apparatus, the data analysis system makes it possible to examine the saved spectra to look for the response of these ions.

We assume that the retention times are close enough so that the elution profiles of the respective molecules overlap or co-elute, but are not exactly coincident. In this case, there is then a moment of time when all three molecules are present in the ionizing source of the MS. Figure 3b shows a spectrum, Spectrum B, collected at that moment in time, when all three ions are present as peaks. Note that each spectral peak is resolved by the MS, meaning there is no overlap.

[FIGURE 3. Three successive spectra, collected in time.]

We know the time the spectrum was collected and thus we know that each of the molecules was eluting from the column at this time. But from a single spectrum alone, it is not possible to determine the precise retention time at which each ion eluted. For example, spectrum B could have been collected from the front of a chromatographic peak, as the molecule began to elute from the column, or from the tail of the chromatographic peak, when the molecule was nearly finished eluting.

We can determine the retention time, or at least the elution order, by examining successive spectra. For example, consider the three successive spectra, A, B, and C in Figure 3, that were collected at successive times, tA, tB, and tC. We can determine the elution order of the respective molecules by examining the relative heights of the peaks as time progresses. As we consider spectra A, B, and C in turn, we see that ion 2 is *decreasing* in intensity relative to ion 1, and we see that ion 3 is *increasing* in intensity relative to ion 1 as time progresses. It follows that ion 2 elutes before ion 1, and that ion 3 elutes after ion 1.

We can confirm the elution order by the following procedure. First, from Figure 3, we obtain the m/z value at the apex of each peak. Given these three m/z values, the DAS extracts from each spectrum the intensity obtained at m/z and plot it versus elution time. Figure 4 plots the three resulting curves, which are, of course, the chromatograms obtained at three values of m/z for ions 1, 2, and 3. As expected, each chromatogram contains a single peak. We immediately confirm that it is ion 2 that elutes at the earliest time.

[FIGURE 4. Chromatograms for three ions.]

The LC/MS data matrix

This analysis suggests that rather than regard the output of an LC/MS as simply a series of spectra, it makes sense to regard the output as a matrix of intensities. We construct the matrix by considering each spectrum to be a column of the matrix. Once in matrix form, we regard each column as a spectrum collected at time t, and each row as a chromatogram collected at fixed m/z. Any a row-oriented cross-section reveals the chromatographic separation, and any column-oriented cross-section reveals the mass-to-charge spectrum.

Thus the first conceptual step in this method is to regard the complete set of spectra as columns of a matrix of responses.

Once we've assembled the spectra into this matrix form, it becomes immaterial that the LC/MS records the data as successive spectra. The output of an LC/MS separation can be described by a matrix of intensities, where the columns are MS spectra, each described by a scan time; and the rows are chromatograms each described by an m/z channel value.

With the data in matrix form, we can examine the matrix by means of a contour plot, Figure 5. From the plot, it is clear that each ion appears as an island of intensity. It is obvious that there are three ions, and that the elution order is ion 2, followed by ion 1, followed by ion 3. Figure 4 also suggest the important role of the apex location. The location of the apex each ion corresponds to the retention time and m/z value for the ion. The height of the apex above the zero value floor of the contour plot measures the ion's intensity.

[Figure 5. Contour plot of simulated LC/MS data matrix]

The counts or intensities associated with a single ion are contained within an ellipsoidal region. The fwhm of this region in the column direction is the fwhm of the mass peak. The fwhm in the row direction is the fwhm of the chromatographic peak.

Drawn through the contour plot are six lines, each corresponding to a row or column. The three horizontal rows are the three chromatograms corresponding to rows that traverse the apex of the respective peaks. The three vertical lines are a series of time, the center of which corresponds to the apex of ion 2.

The effect of co-elution, confusion and noise

Given the spectra, the challenge is to account for all the ions recorded during an LC/MS experiment and to obtain for each ion accurate values for its retention time, m/z, and intensity.

But before we consider how ion detection and parameter estimation is accomplished in the prior art, and how we intend to improve upon the prior art, we must consider two additional non-idealities present in real data that any such post-acquisition method must take into account.

The first effect comes from the finite width of the peaks both in the spectral as well as in the chromatographic directions. The second is the noise present in the instrument.

Figure 6 shows the contour plot that arises from a fourth ion that has an m/z value somewhat larger than that of ion 2, and a retention time also somewhat larger than its retention time. However the apex of the ion 4 lies within the fwhm in both the spectral and chromatographic directions of the apex of ion 2. As a result, ion 4 is both coeluted in the chromatographic direction and interferes with ion 2 in the spectrometric direction. Figure 7 shows the resulting spectra obtained at times A, B, and C. In all spectra, ion 4 appears as a shoulder to ion 2. Note also that in the contour plot there is no distinct apex associated ion 4.

[Figure 6. Example of coeluted ion in contour plot]

[Figure 7. Example of coeluted ion in extracted spectra]

Another non-ideality is noise that adds to signals. Noise comes in two categories. One is the irreducible thermal or shot noise problem inherent in all detection processes. Counting detectors, such as multi-channel plates, add shot noise. Amplifiers, such as electrometers, add thermal, or Johnson noise. Another category of noise is chemical noise; spurious small molecules are inadvertently caught up in process of separation and ionization. Also, complex samples will contain molecules whose concentrations vary over a wide dynamic range. Such samples may include interfering elements, especially troublesome at low concentration. Both detector and chemical noise inevitably occur at some level. These noise sources combine to establish a baseline noise background against which the detection and quantitation of ions must be made.

Adding numerically generated noise to simulate these effects, we obtain the contour plot in Figure 8 and the spectral and chromatographic cross-sections in Figures 9 and 10. Note that in the contour plot, there are now apices through out the plot; in addition, there are now multiple apices associated with ions 1 and 2. These multiple apices, which lie within the fwhm of the nominal apex locations, are artifacts that are due to noise.

[Figure 8. Example of ions in contour plot with noise]

[Figure 9. Example of ions in extracted spectra with noise]

[Figure 10. Example of ions in extracted chromatograms with noise]

Before turning to the invention that is the subject of this application, we must consider yet one more complication that must be faced.

The mixtures analyzed by LC/MS can be complex. Biological samples, especially, can contain 1000's, 10,000's or more, of molecular species whose ions are potentially detectable. In this case, one might simply start peak detection by picking any time range corresponding to a chromatographic FWHM, signal-average the spectra, and examine the result. The resulting spectrum gives the appearance of containing baseline resolved spectral components, so the signal-averaged chromatogram can be extracted, completing the detection step.

As a further complication, the ions might not be baseline resolved. This can happen even in simple mixtures where there are relatively few ions. Ions can then be fused in both dimensions, chromatographically and spectrally. The above procedures of the prior are,

when applied to these chromatograms might yield values, but the fused nature of the peaks will compromise the accuracy of the results.

Ion detection and parameter estimation

A fundamental goal of the LC/MS method is to account for all ions and to determine accurate and precise measurement of their retention time, m/z, and intensity.

This goal is achieved when each potentially detectable ion is in fact detected, and its primary parameters, retention time, m/z, and intensity are estimated from the data. The secondary observable parameters are the widths of the peak in the chromatographic and spectral directions.

As the simulations showed, the detection of ions can be a challenge. Coelution of molecules and interferences produced by near-coincident values of m/z between two ions may cause ions to be missed, producing false negatives. The presence of noise may cause artifacts to be detected, producing false positives. Once the ion is detected, we need to obtain as precise as possible estimates of its retention time, mass-to-charge ratio, and intensity. Again coelution, interference, and noise can hamper these determinations.

Prior art

In the prior art, one examine a series of extracted spectra and chromatograms in order to locate (i.e., detect) the peaks. One then estimates the m/z for an ion by examining a spectrum that contains that ion, and one estimates the retention time for an ion by examining a chromatogram that contains that ion. The response of an ion can be obtained as the height or area of the peak as seen in either trace.

The two step process of detection and parameter estimation is carried out by (1) the examination of spectra and chromatograms, and by (2) the analysis of the responses in the spectra and chromatogram to determine each peak's retention time, m/z, and intensity.

If there are relatively few ions to be detected, one can form a chromatogram by summing all the responses collected over all m/z values within each spectral scan, and plotting these sums against the scan time. The resulting chromatogram is termed a total-ion-chromatogram (TIC). For simple mixtures, each ion might appear as distinct peak in the TIC. But, since, even in simple mixtures, ions might co-elute, we cannot be sure that each isolated peak seen in the TIC is a unique ion. The next step in the detection process is to pick the apex of one peak, and display the spectrum collected at that time. We now see a series of mass peaks, and we can be reasonably sure that each one represents a single ion. As a check, we can plot the chromatogram by picking a channel corresponding to one peak of interest.

Once one has obtained two orthogonal cross-sections, the single channel chromatogram, and the single scan spectrum, the location of the peak apex in the respective plots gives an ion's retention time and value for m/z.

For more complex mixtures, where most molecules are expected to co-elute, the analyst can sum spectral responses only over a subset of the collected channels, e.g, by restricting

the range of m/z channels that are summed. This summed chromatogram is a guide to the ions that were detected within the restricted m/z range. Again, spectra can be obtained for each chromatographic peak apex, and chromatograms for each spectral peak apex. Multiple summed chromatograms would have to be obtained to identify all ions.

In the case where detector noise obscures peaks, one can signal-average the spectra or signal-average the chromatograms to average out the effects of noise.

Thus, to obtain more precise peak parameters, the analyst can co-add spectra that encompass a chromatographic peak to reduce the effects of noise. The m/z values and areas and heights can be obtained from this averaged spectrum. Analogously, co-adding chromatograms centered on the apex of a spectral peak can produce chromatograms with less noise, providing more precision estimates of retention time and areas and heights.

To obtain the peak parameters, the retention time, m/z and intensity, one typically applies a peak-finding and parameter extraction algorithm to each extracted or averaged trace. In the prior art, given a spectrum, one applied a centroiding or peak detection algorithm to this spectrum to extract estimates of intensity and m/z.

From the chromatographic trace, the algorithm will extract the retention time and the peak height or area. From the spectral trace, the algorithm will extract the mass-to-charge ratio and the peak height or area. Such algorithms typically take as input points on the up and down slopes of the respective peaks and combine these points using one or another fitting routine. For example, a quadratic, parabolic fit might be applied to the top few points in each peak in order to obtain a precise estimate of its apex location.

The procedure applied above produces four estimates of the intensity of the peak. These are the area and height for each of the two traces. It is up to the analyst to decide which value, or combinations of values, to adopt in performing quantitative work.

Once an ion is detected by this manner, algorithms, known as peak-detection algorithms can be applied to either or both curves to estimate parameters. The apex location of the chromatogram gives the retention time. The apex location of the spectrum determines the value for m/z. The response can be obtained by determining either the area or heights of the peaks in either representation.

Summary of Prior Art

A common method of the prior art used to detect ions is to form a total ion chromatogram, or subsets of total ion chromatogram. The analyst can select a time range that encompasses the fwhm of each peak and form the m/z spectrum by signal-averaging these spectra over the fwhm of each peak. One determines the m/z and height or area for each ion by applying a peak detection algorithm to the resulting spectra. This method will detect ions that elute at a common retention time, provided the MS resolves each ion.

One can confirm the identification by selecting a range of m/z channels that encompasses the fwhm of each spectral peak and form the chromatogram by signal-averaging these channels. One then identifies determines the retention time and height or area for each peak, by applying a peak detection algorithm to the resulting chromatogram.

The analyst can choose the height or area of ions from either the spectrum or from the chromatogram as possible measures of intensity.

Problems with the Prior Art

There are several problems with these procedures of the prior art that make it difficult to reliably detect ions and estimate their parameters.

The detection methods are tedious if carried out manually and somewhat subjective if carried out manually or automatically.

It is not possible to obtain the most accurate value for m/z from a single extracted spectrum. It is not possible to obtain the most accurate value for retention time from a single extracted chromatogram.

But the simple signal-averaging schemes described do not produces estimates of retention time, m/z, or intensity that have the highest possible precision, or lowest possible statistical variance, given the data. There is no clear rule as to how many chromatograms to co-add; there is no clear rule as to how many spectra to co-add. Including too many may cause the analyst to combine peaks; including too few may not reduce noise in an optimal fashion.

These procedures will not be guaranteed to give uniform reproducible results for ions at low concentration, or for complex chromatograms, where coelution and ion-interference may be a common problem.

Invention: Novel method to detect and quantify ions in LC/MS

The method disclosed here accomplishes the detection and quantification of ions from peaks found in an LC/MS chromatogram with a novel method.

Summary

If the data matrix is free noise and if none of the ions interfere, then each ion produces a unique, isolated island of intensity. Concentric contours identify each island, and the inner-most contour within an each island identifies the element having the highest intensity. That element is a local maximum of intensity, meaning that its intensity is greater than that of its immediate eight neighboring elements. We will refer to such as element as a maximal element, or simply as a local maximum.

Figure 5 showed how each island contains a single maximal element. The uniqueness of this maximal element then suggests a simple detection method: interrogate each element in the data matrix, identify all elements that are local maxima of intensity, and label each such local maximum as an ion. We then obtain the parameters of the ion from the maximal element. The ion's retention time is the time of the scan containing the maximal element; its m/z is the m/z for the channel containing the maximal element; and its intensity is the intensity of the maximal element itself.

But this detection and quantification method is not adequate for several reasons. The presence of noise means that many local maxima will be due to noise, not ions. Even if a threshold criterion is applied to the ion's intensity to reduce these false positives, Figure 8 shows that noise might produce more than one multiple local maxima for an ion. Thus ions could be double counted. Further, Figure 6 showed that a pair of ions that co-elute in time and interfere spectrally may produce only a single local maximum, not two. Thus an ion appearing in the data matrix with significant intensity might be missed.

Finally, this simple method is not a statistically optimum method. The variance in the estimates of ret time, m/z and intensity are determined by the noise properties of a single element; the method does not make use of the other elements in the island of intensities surrounding the maximal element to reduce variance in the estimate.

For a single-channel of data, a known way to reduce the effects of noise is by smoothing. Smoothing can be carried out by convolving the data array with a set of fixed-value filter coefficients. The coefficients of FIR filters can be chosen to carry out a variety of task include smoothing and differentiation. Well-known filters that can be used to smooth or differentiate one-dimensional arrays of data are the Savitzky-Golay filter. Such filter methods are well-known and are also referred to as finite-impulse response (FIR) filters.

The LC/MS data matrix is an example of a two-dimensional array, where the dimensions are time and m/z. Such an array can be modified by convolving the data matrix with a two-dimensional array of filter coefficients. For example, elements of the convolution matrix can be chosen to correspond to Savitzky-Golay smoothing or differentiation filters, among other filter shapes. The filter coefficients can be chosen to perform smoothing or differentiation operations on the underlying data matrix.

We can now summarize the preferred method disclosed here. The first step is to choose a size for the two-dimensional convolution matrix and values for its filter coefficients. Given the convolution matrix, the second step is to apply them to the LC/MS data matrix, using the rules of matrix convolution, thereby obtaining a convolved data matrix. Figure 10.1 shows the LC/MS data matrix after convolution with a smoothing filter.

[Figure 10.1 Ions after convolution]

The third step is to find all local maxima in the convolved data matrix, by the method described above. We then determine and apply a threshold, retaining only those local maxima whose (filtered) intensities lied above that threshold. Each retained local maximum is identified as an ion. This completes the detection of the ions under this method.

The parameters of the ion, its retention time, m/z, and intensity are obtained from the elements of the convolved data matrix. We could proceed in analogy to the method described above, where the ion's retention time is now the time of the (filtered) scan containing the (filtered) maximal element; its m/z is the m/z for the (filtered) channel containing the (filtered) maximal element; and its intensity is the intensity of the (filtered) maximal element itself.

However, the preferred method is to fit a parabola to the elements of the convolved data matrix that surround the maximal element of the convolved matrix that identify an ion. A parabola is a good approximation to the shape of the convolved peak near its apex. The method uses a parabolic fit in order to find an interpolated value for the ions parameters. An interpolated value will provide more accurate estimates of retention time, m/z and intensity than those obtained by reading of values of scan times and spectral channels.

Thus, we obtain the ion's retention time by fitting a 2-dimensional parabola to the maximal element and its eight surrounding neighbors in the convolved data matrix. The fitting procedure is preferably implemented with a linear-least-square optimization. The ion's retention time is now the time of the maximum of the interpolated parabola; its m/z is the m/z at the maximum of the parabola; and its intensity is the intensity at the maximum of the parabola. The final step is to collect these results into a tabular list. For example, each row in the list is an ion. The first column contains the ion's retention time, the second column contains its mass-to-charge ratio, and the third column contains its intensity. The output of the method is a list of ions. This list is then input to further well-known operations, which will be described below.

This method is summarized in Figure 12.

FIGURE 12. The ion detection and parameter estimation method

One of the key steps in this method is the determination of the size of the convolution matrix and the values for the filter coefficients. Possible methods for making this choice, and the preferred method, will be described in detail below, as well as the advantages to using one set of coefficients over another. Also described below is the Matched Filter Theorem which shows how to determine the signal-to-noise ratio that results from a convolution, and also shows how to compute filter coefficients that maximize the signal-to-noise ratio.

A second key step is the method for finding a threshold. Again, possible methods and the preferred method for making this choice will be described in detail below.

Advantages over prior art

The method proposed here, which is convolution, followed by apex detection, threshold selection, and parameter estimation, is an improvement over the prior art methods described above.

The convolution operation is a more general and powerful approach than the simple signal-averaging schemes of the prior art. The values for the convolution coefficients can be chosen to obtain values for retention time, m/z, and intensity, with signal-to-noise ratios that are enhanced over what can be obtained from the extraction of single channels or scans. They can be chosen to produce estimates of retention time, m/z, and intensity that have the highest possible precision, or lowest possible statistical variance, given the data. As a result, the proposed method gives more reproducible results for ions at low concentration.

The coefficients of the convolution matrix can be chosen to resolve ions that are co-eluted and interfering. The apices of shouldered ions can be detected, thus addressing the limitations of the prior art for the analysis of complex chromatograms, where coelution and ion-interference may be a common problem.

The convolution operation itself is linear and non-iterative. The preferred method of implementation is by means of a general purpose programming languages implemented in a general purpose computer. It is also possible to implement convolution in special purpose processors, known as digital-signal-processors (DSP) that provide enhanced processing speed. The identification of ions as local maxima within the convolved matrix is an automatic, objective, and rapid operation.

Graphical summary of method

Figure 12 summarizes the method of detecting ions and establishing their parameters.

Figure 13 summarizes the application of the threshold to the ion parameter list.

[FIGURE 13 Threshold applied to ion list]

Convolution and the Matched Filter

The method disclosed describes the application of the mathematical operation of convolution to a matrix of intensities. The convolution operation has the effect of combining two input matrices to form one output matrix. The result of this operation, the output, is a matrix of convolved intensities.

A key step in the method is the convolution of the two-dimensional LC/MS data matrix with a filter matrix. It will simplify the discussion to first define convolution for the one-dimensional case and then to introduce the Matched Filter Theorem (MFT) in this context. The next section will then generalize convolution and the MFT to the two-dimensional case. The filter found by the MFT will guide the choice of convolution filters used by the proposed method.

The convolution operation is a linear operation. It is also a non-iterative, open-loop operation. It is with no prior knowledge of the location, number of intensity of any of the ions in the LC/MS chromatogram. The convolution operation can provides a statistically optimum averaging of each of the components in the LC/MS chromatogram.

Convolution of one-dimensional data

Convolution is a linear operation that combines two input arrays to produce an output array. We regard one of the input arrays as a data array that can vary from experiment to experiment; the other array is a set of fixed filter coefficients. We convolve the input array with the filter array to obtain the output array. For specificity, we will assume the one-dimensional array is a chromatographic trace, and the each array elements represents a successive sample time.

Given a one-dimensional, N-element, input array of intensities d_i and the filter coefficients f_i we define their convolution as

$$c_i = \sum_{j=-h}^{h} f_j d_{i-j}$$

where c_i is the output, convolved array.

The filter array f_j contains M elements. We assume that M is an odd number for convenience. The index j goes from $j = -h,\ldots,0,\ldots h$, where we have defined $h \equiv (M-1)/2$. The value of c_i then corresponds to a weighted sum of the h elements that surround d_i.

Typically, we have that $M \ll N$. Spectra and chromatograms are examples of one-dimensional input arrays that contain peaks. The width of the peaks set the width of the convolution filters. The peaks have widths $(\approx M)$ much smaller than the length N of the input array.

The index for d_i ranges from 1 to *N*. But note that c_i is defined only for $i \geq h$ or $i \leq (N-h)$. The value for c_i near the array boundaries, when $i < h$ or $i > (N-h)$, are not defined by the summation. We can handle these edge effects by simply limiting the values for c_i to be those where the summation is defined; the summation then applies only to those peaks far enough away from the array edges so that the filter f_j can be applied to all points within the peak. Generally, this is not a significant limitation of the method.

The Matched Filter Theorem for one-dimensional data

The coefficients for f_j are often chosen to produce a smoothing or differentiation function. But our goal is to find coefficients for f_j that perform a *detection* function. The Matched Filter Theorem (MFT) justifies use of convolution as part of a detection method. The MFT assumes that the data array d_i can be modeled as a sum of a signal $r_o s_i$ plus additive noise, n_i.

$$d_i = r_o s_{i-i_o} + n_i .$$

The shape of the signal is fixed and described by a set of coefficients, s_i; the scale factor r_o determines the signal amplitude. The MFT also assumes that the signal is bounded, that is, it is zero (or small enough to be ignored) outside some region. We assume that signal extends over *M* elements; for convenience, we assume that the *M* is odd and that center of the signal is at s_o. If we define $h \equiv (M-1)/2$, then $s_i = 0$ for $i < -h$ and for $i > h$. In the above expression, the center of the signal will appear at $i = i_o$.

For the noise, we will only consider the simple case where the n_i .are assumed to be uncorrelated Gaussian deviates, with zero mean and a standard deviation of σ_o; more general formulations for the MFT accommodate correlated or colored noise. We will consider the case of Poisson noise later.

The signal-to-noise (SNR) of each element is then $r_o s_i / \sigma_o$. What is the SNR of a weighted sum of the data that contains the signal s_i? Consider an M-element set of weights w_i, where $h \equiv (M-1)/2$, and $i = -h, \ldots, 0, \ldots h$. We center the weights to coincide with the signal, and we define the weighted sum *S* as

$$S = \sum_{i=-h}^{h} w_i d_{i-i_o} = r_o \sum_{i=-h}^{h} w_i s_i + \sum_{i=-h}^{h} w_i n_{i-i_o} .$$

To compute signal-to-noise ratios, we need to consider the statistical properties of the sum *S*. Consider an ensemble of arrays, where the signal in each array is the same, but the noise is different. The average value of *S* over the ensemble is

$$\langle S \rangle = r_o \sum_{i=-h}^{h} w_i s_i ,$$

because in an ensemble average, the noise term has a mean value of zero, so the noise term drops out.

We now apply the weight to a region containing only noise. The ensemble mean of the sum is zero. But the standard deviation of the weighted sum about the ensemble mean is

$$\sigma \equiv \sqrt{\left\langle \left(S - \langle S \rangle \right)^2 \right\rangle} = \sigma_o \sqrt{\sum_{i=-h}^{h} w_i^2} .$$

Finally, we compute the SNR as

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\left(\sum_{i=-h}^{h} w_i s_i \right)}{\sqrt{\sum_{i=-h}^{h} w_i^2}}$$

This result is for a general set of filter coefficients w_i.

The MFT provides an answer to the question: what values for w_i will maximize the SNR? We can see the answer immediately, if we regard the weighting factors as elements of an M dimensional vector w of unit length. That is we assume the weighting factors are normalized so that $\sqrt{\sum_{i=-h}^{h} w_i^2} = 1$, then it is clear then that the ratio is maximized when the vector w points in the same direction as the vector s. The vectors point in the same direction when respective elements are proportional to each other, or when $w_i \propto s_i$. The Matched Filter Theorem says that the weighted sum has the highest signal-to-noise when the weighting function is the shape of the signal itself.

If we set $w_i = s_i$, so for noise with unit standard deviation, we have that

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\left(\sum_{i=-h}^{h} s_i^2 \right)}{\sqrt{\sum_{i=-h}^{h} s_i^2}} = \frac{r_o}{\sigma_o} \sqrt{\sum_{i=-h}^{h} s_i^2}$$

We have now established the signal properties of the weighted sum when the filter coefficients are centered on the signal, and we have established the noise properties when the filter is in the noise-only region.

The MFT makes it clear how the convolution operation can be used to detect a signal. The convolution operation consists of moving the filter coefficients down the data array and obtaining a weighted sum at every point. Assume the filter that is matched to the signal $w_i = s_i$, satisfying the MFT. In the noise-only region of the data, the amplitude of

the output will be dictated by the noise. As the filter overlaps the signal, the amplitude will increase, and reach a unique maximum when the filter is aligned in time with the signal.

Convolution of two-dimensional data

The matrix of intensities produced by the LC/MS experiment is the input to the two-dimensional convolution. To obtain the output matrix , the LC/MS data matrix is convolved with a matrix of filter coefficients. The output matrix has essentially the same number of rows and column elements as the input LC/MS matrix.

For specificity we assume the LC/MS matrix is rectangular and that the size of the matrix of filter coefficients is comparable to the size of a peak. That is the size coefficient matrix is much smaller than of the input data matrix or output convolved matrix.

An element of the output matrix is obtained from the input LC/MS matrix as follows: the filter matrix is centered on each input element, and then the filter elements multiple the corresponding data elements and the products are summed. The procedure is described algebraically as follows.

It is straightforward to generalize the convolution operation to the case of two-dimensional data. The input now consists of an array $d_{i,j}$ subscripted by two indices, i, j, where $i = 1,\ldots,M$ and $j = 1,\ldots,N$, and, again, the array's values can vary from experiment to experiment. The other input array is a set of fixed filter coefficients, $f_{p,q}$, also subscripted by two indices. The filter coefficients, $f_{p,q}$, is a matrix that has $P \times Q$ coefficients. We define $h = (P-1)/2$ and $l = (Q-1)/2$, so we have $p = -h,\ldots,h$, and $q = -l,\ldots,l$.

The convolution of $d_{i,j}$ with $f_{p,q}$ gives the output array $c_{i,j}$ where,

$$c_{i,j} = \sum_{p=-h}^{h} \sum_{q=-l}^{l} f_{p,q} d_{i-p,j-q} .$$

Generally, the size of the filter is much less than the size of the data matrix, so that $P \ll M$ and $Q \ll N$. The above equation says that we compute $c_{i,j}$ by centering $f_{p,q}$ on the (*i,j*) th element of $d_{i,j}$ and then using the filter coefficients $f_{p,q}$ to obtain the weighed sum of the surrounding intensities.

Thus each element of the output matrix $c_{i,j}$, obtained by the convolution operation, corresponds to a weighted sum of elements of $d_{i,j}$. Each element $d_{i,j}$ is obtained from a region centered on the i,jth element.

We acknowledge and ignore edge effects…

The matched filter for a two-dimensional peak

The MFT immediately generalizes to the case of a bounded, two-dimensional signal embedded in a two-dimensional array of data. As before, we assume the data is modeled as a sum of signal plus noise.

$$d_{i,j} = r_o s_{i-i_o,j-j_o} + n_{i,j}$$

where the signal $S_{i,j}$ is limited in extent and whose center is located at (i_o, j_o) with amplitude r_o. Each noise element $n_{i,j}$ is an independent Gaussian deviate of zero mean and standard deviation σ_o.

What is the SNR of a weighted sum of the data that contains the signal $S_{i,j}$? Consider an $P \times Q$-element set of weights $w_{i,j}$, where $h = (P-1)/2$ and $l = (Q-1)/2$, so we have $p = -h,\ldots,h$, and $q = -l,\ldots,l$. We center the weights to coincide with the signal, and we define the weighted sum S as

$$S = \sum_{i=-h}^{h} \sum w_i d_{i-i_o,j-j_o} = r_o \sum_{i=-h}^{h} \sum w_{i,j} s_{i,j} + \sum_{i=-h}^{h} \sum w_{i,j} n_{i-i_o,j-j_o} .$$

The average value of S over the ensemble is

$$\langle S \rangle = r_o \sum_{i=-h}^{h} \sum w_{i,j} s_{i,j} ,$$

and the standard deviation of the noise is

$$\sigma = \sigma_o \sqrt{\sum_{i=-h}^{h} \sum w_{i,j}^2} .$$

and the signal-to-noise ratio is then

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\sum_{i=-h}^{h} \sum w_{i,j} s_{i,j}}{\sqrt{\sum_{i=-h}^{h} \sum w_{i,j}^2}}$$

In parallel to the logic for the one-dimensional case, the value for the signal-to-noise is maximized when the shape of the weight function is proportional to the signal, that is when $w_{i,j} \propto s_{i,j}$

We have now established the signal properties of the weighted sum when the filter coefficients are centered on the signal, and we have established the noise properties when the filter is in the noise-only region.

The MFT makes it clear how the convolution operation can be used to detect a signal. The convolution operation consists of moving the filter coefficients down the data array

and obtaining a weighted sum at every point. Assume the filter that is matched to the signal $w_i = s_i$, satisfying the MFT. In the noise-only region of the data, the amplitude of the output will be dictated by the noise. As the filter overlaps the signal, the amplitude will increase, and reach a unique maximum when the filter is aligned in time with the signal.

If these conditions hold, then the MFT defines a detection method as follows: Choose filter coefficients to correspond to the (assumed known) shape of the underlying signal. Convolve the data with that shape. Identify the highest value in the convolved data, and check that its SNR is above the predetermined detection threshold. If the filtered response meets or exceeds the threshold, then the signal is detected; the arrival time and amplitude of the signal is given by the time and value of the local maximum in the column direction, and the best fit estimate of m/z is obtained from the row direction.

Apex Location and peak detection

In the detection method disclosed here, the step of detecting ions is performed on the elements of the convolved data matrix $c_{i,j}$.

The presence of an ion produces a peak, with a characteristic local maximum, in the convolved intensity. It follows that any local maximum in the convolved output is then a candidate for being a peak. In the absence of detector noise, *every* local maximum would identify the presence of a peak and its corresponding ion. But in presence of noise, many low-amplitude local maxima will be due to noise and will not correspond to a genuine peak. Individual noise artifacts give rise to local maxima in the convolved matrix.The method disclosed here accepts a local maximum as a peak only if the amplitude of that local maximum is above a threshold value. This threshold value is set to make it highly unlikely that a local maximum that equals exceeds that threshold is due to noise.

Thus after the step of local maximum detection, the next step in the method is to pick a suitable threshold and to compare the amplitude of each local maximum to that threshold. The method identifies only those local maxima whose convolved amplitude exceeds the threshold as a detected peak.

Apex detection

Each ion produces a unique apex in the matrix of convolved intensities. It is the locations of the unique maxima in the convolved matrix that gives us information on the number, and properties of the ions present in the sample.

Thus, after convolution, the next step of the method is to identify all the local maxima of the convolved data. For one-dimensional data, a local maximum is any point whose amplitude is greater than its two nearest neighbors. For two-dimensional data, cross-sections through the apex yield a bell-shaped curve with a single maximum. Looked at from above, with a contour plot, each peak correspond to a single maximum, or apex Thus, for two-dimensional data, a local maximum or apex is any point whose amplitude is greater than its eight nearest-neighbor elements. For example in the following matrix,

the central element is a local maximum because all adjoining elements have value less than 10.

8.5	9.2	6.8
9.2	10.0	8.4
7.9	8.5	7.2

Detection threshold

We declare a local maximum to be an ion if its value is above a threshold.

The value of the detection threshold can be obtained by subjective or objective means. Regardless of how the value is arrived at, the effect of the detection threshold to divide the distribution of true peaks into two classes: those that are above the threshold and those that are below the threshold. The true peaks below the threshold are false negatives and are missed by the method. The threshold also divides the distribution of noise peaks into two classes, those which are above the threshold and those below the threshold. The noise peaks above the threshold are termed false positives.

In common practice, the detection threshold is set according to a desired false positive rate. That is the threshold is set to the chance that a noise peak will equal or exceed the threshold in a given experiment is highly unlikely. Many practitioners term the chance that a given peak above the threshold is in fact due to noise, the confidence level.

To obtain fewer false positives one sets the detection threshold to a higher value. A lower false positive rate means a somewhat higher false negative rate; i.e., low-amplitude, genuine peaks will not be detected.

An example of a subjective method is to simply draw a line that is close to the maximum of the observed noise. Henceforth, all local maxima above this threshold are peaks. All local maxima below the threshold are noise.

The preferred method for setting the threshold for convolved data is to use an objective method based upon a histogram of the data. Figure xx shows a histogram of the data. The true peak is identified as an outlier and is ignored. The standard deviation of the distribution is obtained by conventional means. Two examples of threshold are set. One corresponds to 2 standard deviations. One corresponds to 4 standard deviations.

A variation of the empirical method can make use of the fact that the standard deviation σ of the convolved output noise is related to the standard deviation σ_o of the input noise as given by $\sigma = \sigma_o \sqrt{\sum_{i=-h}^{h} s_i^2}$. This formula assumes that the input noise is uncorrelated Gaussian deviates. Thus the input noise could be measured and the standard deviation of the output can be inferred, knowing only the values used for the filter coefficients.

Note that the goal of any method is to simply determine an intensity value, the threshold, which is then used to edit the ion list. All ions whose intensities are below the threshold

are considered noise. They are rejected and are not included in further analysis. Regardless of which method is used, the effect of either method is to simply edit the list. No modifications are made to the values of the retained ions.

The Figure 13 illustrates these methods for determining and applying a threshold.

[FIGURE 13 Threshold applied to ion list]

Method to determine peak parameters

After the method identifies which local maxima are peaks, the next step is to estimate the parameters for each peak. These parameters are the retention time, mass-to-charge ratio, and intensity. Additional parameters are the chromatographic peak width and the mass-to-charge peak width.

A parameter estimation method has to contend with the fact that the elements of the convolved matrix digitally sample the data. As a result of this discrete sampling, the apex of a peak in time may not coincide exactly with a sample time; and the apex of a peak in mass-to-charge may not coincide exactly with an m/z channel.

In general the actual maximum of the signal in time and mass-to-charge will be offset from by a fraction of the sample period or the mass-to-charge channel interval. Thes fractional offsets can be estimated from the values of the matrix elements surrounding the apex element..

In the case of one-dimensional data, the preferred method is to first to locate the element of c_i that is the maximum, and then locate the two adjoining elements. A parabola is fit to the intensities at these three points; it is the maximum of that fitted parabola that locates the time amplitude of the maximum value of the convolution. The time of the maximum of the parabola is the best estimate of the arrival time of the peak. Both the amplitude and the arrival time obtained from this fitting procedure are optimum estimates. The Nyquist sampling theorem is the formal justification of the interpolation process.

Note that the effect of the convolution is to combine the data in the bulk of the peak so that all the information about the signal's sample and arrival time has been compressed into the local maximum. It is only the highest element in the convolved response that contains the information we need about the signal.

For two-dimensional data, the preferred method is to take the matrix of 9 values and fit a two-dimensional parabolic shape. The value of the parabola at the maximum, and its interpolated x and y values become the estimates of ion intensity, m/z, and retention time.

To summarize: the precise, interpolated location of maximum gives, in the row direction an optimum estimate of retention time; the precise, interpolated location of the maximum in the column direction gives an optimum estimate of mass-to-charge ratio. The precise height of the apex above baseline gives an optimum estimate (scaled by filter factors) of ion intensity or concentration.

Example of a one-dimensional Gaussian Matched Filter

To make these results concrete, consider the case where the signal is a single peak. We model the peak as a Gaussian whose width is given by the standard deviation σ_p, where the width is measured in units of sample elements. The signal is then

$$r_i = r_o \exp\left(-\frac{1}{2}\frac{(i-i_o)^2}{\sigma_p^2}\right)$$

Assume that we set a boundary for the filter to correspond to $\pm 4\sigma_p$. It will be useful to compare the signal-to-noise properties of two filters. One is the matched filter, which is just the signal shape itself, centered on zero, and bounded by $\pm 4\sigma_p$:

$$f_i = \exp\left(-\frac{1}{2}\frac{i^2}{\sigma_p^2}\right), \quad \text{for } i > -4\sigma_p \text{ and } i < -4\sigma_p .$$

The other filter is a simple running average, or box car filter, where

$$f_i = \frac{1}{M} = \frac{1}{8\sigma_p + 1}.$$

The output of this filter is an average value over M points.

To consider a particular example, we will assume that the system samples four points per standard deviation, so we can set $\sigma_p = 4$ and consider a filter that is 33 points wide. For a Gaussian peak of unit height, the running average filter gives the average signal over the peak to be $0.304\, r_o$, and the standard deviation of the noise is $\sigma_o / \sqrt{33} = 0.174\, \sigma_o$, for an SNR of $1.75(r_o / \sigma_o)$.

For the matched filter, we have that the maximum signal is $7.09\, r_o$, and the noise amplitude is $2.66\, \sigma_o$ for an SNR of $2.66(r_o / \sigma_o)$. Thus the matched filter produces an SNR that is over 50% higher than that provided by the simple running average.

The running average filter has a profile of that has a constant value. Such filters are called boxcar filter. The convolution of a boxcar filter with the Gaussian peak shape will still produce an output that has a unique maximum value. Thus either of these filters can be used in the convolution step that is the heart of this method. Both are linear and both will produce a unique maximum. The matched filter profile is preferred because of the higher SNR at the local maximum.

Example of a two-dimensional Gaussian Matched Filter

TBD

Summary

The MFT assumes that the data is the sum of a signal plus noise. The signal is bounded and occurs at an unknown time and with unknown amplitude. The noise points are Gaussian deviates, with zero mean and uniform standard deviation.

The MFT shows that the weighting coefficients that maximize the SNR of a weighted sum are those that describe the signal itself. It follows that the operation of convolution, followed by the detection of the maximum is an optimum method by which to detect the presence of the signal. The interpolated height and position of the apex is the optimum estimate for the amplitude and arrival time of the signal.

Knowing the standard deviation of the noise and the filter coefficients, we can obtain the standard deviation of the maximum value. Thus from this value we can determine the likelihood that it is due to noise. We can set a threshold value based upon an acceptance rate of false positives.

Linear weighting coefficients other than those that follow the signal shape are possible, and though they may not produce the highest possible SNR, they may have other counter-balancing advantages.

The MFT suggests the detection method we employ: Choose filter coefficients based upon the shape of the underlying signal. Convolve the data with that filter. Identify the highest value in the convolved data, and check that its SNR is above the predetermined detection threshold. If the filtered response meets or exceeds the threshold, then the signal is detected; the arrival time and amplitude of the signal is given by the time and value of the interpolated local maximum.

Properties of the ion parameters

The ion detection and quantitation method results in three fundamental measurements of each ion. These are the ion's retention time, m/z, and a measurement of intensity. The measurement of intensity is simply the response of the filter output at the local maximum. Note that the intensity measurement does not correspond to the peak area or the peak height. However, the intensity measurement must be in proportion to those values, since the convolution operation produces a linear combination of intensity measurements.

Measures of intensity

The set of convolution coefficients will determine the scaling of the intensity. The next section will describe the possible and preferred convolution coefficients. Each set will give a different intensity scaling. Regardless of the details of intensity scaling, as long as a consistent set of filters is used to determine the intensities of standards and calibrators and sample, the resulting intensity measurements will give accurate, quantifiable results. That is, the disclosed ion detection and parameter estimation method will produce results suitable for the quantitative analysis of components. The intensities can be used to establish concentration calibration curves, and the concentration of analytes can be obtain from the results of this method.

Measures of peak width

In addition to intensity, other ion properties can be obtained from the convolution results. These properties include, but are not limited to, the widths of the ion peaks. Each convolved peak still has a width in both the chromatographic and the spectral directions. Conventional means of measuring these widths can be applied to the convolved peak corresponding to each ion detected in the convolved data matrix. The following section will describe two types of filters that can be applied in either the chromatographic or the spectral directions. If a smoothing filter is applied, the peak width corresponds to the fwhm in that direction. If a deconvolving second derivative filter is employed, the appropriate measure of peak width is the width between zero-crossing points, as will be described. Thus, in addition to retention time, m/z, and intensities, two measures of peak width can be obtained form each ion, for a total of five possible measurements per ion.

Measures of error

Each of the five measurements obtained for each ion has an error associated with it. The errors can only be known in a statistical sense. As is true for any measurement, the associated error is made up of two distinct contributions. One contribution is a systematic or calibration error. For example, if the MS m/z axis is not perfectly calibrated, than any given m/z value will contain an, essentially identical, offset. Such an error is independent of the signal-to-noise or amplitude of a particular ion. In the case of m/z, this error is independent of the m/z peak width.

The second contribution to error is the irreducible statistical error associated with each measurement. This origin of this error resides in either in thermal or shot-noise related effects. The magnitude or variance of this error for a given ion depends on the ion's peak width and intensity. It is a measure of reproducibility, and is therefore independent of calibration error. Another term for statistical error is precision.

The statistical error associated with each measurement can in principle be estimated from the fundamental operating parameters of an instrument. For example in a MS analyzer these would be the ionization and transfer efficiency of the instrument coupled with the efficient of the micro-channel counting plate (MCP) which all together determine the counts associated with an ion. Via Poisson statistics, the counts determine the statistical error associated with any of the five given measurements. Each error can be obtained from counting statistics via the theory of error propagation.

In practice, statistical errors can be inferred from the data directly. This can be accomplished by investigating the reproducibility of measurements. For example replicate injections of the same mixture can establish the statistical reproducibility of values of m/z of the same molecules. The statistical reproducibility of retention time measurements is more difficult to accomplish. This is because the systematic errors that arise from replicate injections generally mask the statistical error. However samples may well contain molecules that when ionized may produce ions at different values for m/z. Since these ions originate from a common molecule, then in an LC/MS system, the intrinsic retention time of each such ion must be identical. The only difference between

measurements of the retention times of such molecules are statistical errors associated with the fundamental detector noise associated with measurements of peak properties.

Thus measurements of the retention time differences between ions that come from the same molecule within an injection can measure the statistical error associated with an ions' retention time.

The errors associated with intensity of an ion are problematic in they arise from a combination of statistical and systematic effects associated with the ionization (systematic error) as well as detection (statistical error). Again, the statistical effect may be isolated by comparing ratios of intensities from ions that arise from a common parent within an injection.

In principle then, each of the five measurements can be accompanied by measures of the statistical and systematic errors associated with each measure. Thought these errors apply to each individual ion, their value can generally be inferred by analyzing sets of ions. This is in contrast to the five measurements that are unique to each in. After a suitable analysis, the errors associated with each measurement can be included in each row of the table, yielding then a total of fifteen measurements that can be associated with each ion; there are the five measurements and each of their statistical and systematic errors.

Statistical error of retention time and m/z measurement

The statistical component of error, or precision, in retention time and m/z depends on the respective peaks widths and intensities. For a peak that has a high SNR, the precision can be substantially less than the fwhm of the respective peak widths. In this section the fwhm is the that of the peak in the LC/MS chromatogram, before convolution (and not the fwhm of the convolved peak).

For example, for a peak that has a fwhm of 20 milli-amu and high SNR, the precision can be less than 1 milli-amu. For a peak that is barely detectable above the noise, the precision can be 20 milli-amu.

The analysis of the relationship between peak width, convolution filter coefficients and signal-to-noise of the peak shows that precision is proportional to the peak width and inversely proportional to peak amplitude. The general result can be expressed as

$$\sigma_m = k \frac{w_m}{h_p}.$$

In this expression, σ_m is the precision of the measurement of m/z (expressed as a standard error), w_m is the width of the peak (expressed in milli-amu at the fwhm), h_p, is the intensity of the peak, expressed as a post-filtered, signal to noise ratio, and k is a dimensionless constant of order unity. The exact value for k depends on the filter method used.

This expression shows that σ_m will be less than w_m, the fwhm of the peak.

The same argument applies to the measurement of retention time. The precision to which we can measure retention time of a peak depends on the combination of peak width and

signal intensity. If the fwhm max of the peak is 0.5 minutes, we can measure the retention time to a precision, described by a standard error, of 0.05 or 3 seconds.

Advantages of method

One of the advantages of the method of detection and parameter estimation is that it lends itself to the determination of each of the measures cited above for each ion. In particular, this method measures the retention time and m/z of each ion with a statistical error or precision that is less than the fwhm of the respective peak widths as measured in the original LC/MS data.

Filter coefficients for performing convolution

This section returns to the filters that can be employed in the convolution step of the ion detection and parameter estimation method.

For a Gaussian peak, the Matched Filter Theorem (MFT) specifies the Matched Gaussian Filter (MGF) as the filter whose response has the highest signal-to-noise ratio as compared to any other convolution filter. But the key characteristic that is central to the method is that for an individual ion the convolved output is a peak with a single local maximum. It is the number and location of each local maximum that specifies the number and the properties of each ion.

Thus the only requirement the method places on a convolution filter is that its output have a unique maximum. Since the input signal—the peak in the LC/MS data matrix has a unique maximum, the requirement placed on the convolution filter is simply that it, too, have a unique positive maximum. For an ion that has a bell-shaped response, this condition is satisfied by any convolution function whose cross sections are all bell-shaped, with a single positive maximum.

Specifically, any convolution filter that has the property that it has a unique, positive valued apex makes that filter suitable to be used for the method. A contour plot of the filter coefficients reveals the number and location of the local maxima. All row and column and diagonal cross section through the filter must have a single, positive, local maximum.

There are a large number of filters shapes than can be employed. Examples of suitable filters are those whose cross-sections are bell-shaped, such as inverted parabolas, triangle filters, or co-sinusoids. Even a filter that has a constant value (a box car filter) is suitable, since its convolution with a peak will produce an output that has a single maximum. The widths of these filters can be matched to the fwhm of the peak (in time and in mass-to-charge), but that is not a necessary requirement.

Thus, for example, a Gaussian filter that has widths other than the fwhm of the ion can be used. In fact, any filter whose cross section has a single, positive, local maximum is a filter that is suitable for the method proposed here.

Other possible filter cross-sections are those that have a single, positive local maximum, but have negative side-lobes. Filters that extract second derivatives, or curvature have these characteristics. The coefficient values for second derivative filters sum to zero, and this requirement is compatible with the method proposed here.

Details of filters

A suitable smoothing filter is generally a symmetric, bell shaped curve, with all positive values, and a single maximum. Savitzky-Golay polynomial filter provide a family of smoothing filters. The 0th order filter is a flat top, box car filter. The 2nd order filter is a parabola that has a single, positive maximum.

Asymmetric, tailed curves are possible, but don't confer much advantage over a symmetric bell shape.

Examples of smoothing filters are: Gaussian shapes, triangle shapes, and parabolas, all with single maxima. A flat top, box car shape is possible too. It does not have a unique maximum value. But when convolved with a bell-shaped peak will produce a single maximum.

A flat-top, or box car shape produces a minimum variance for a given number of filter points, a fact that is well-known. But it is also well known that it has a poor transfer function. That is, it passes high frequency noise. Thus double counting can result at low amplitude as a result of convolution with baseline noise. A flat-top, box car can be implemented with fewer multiplications than a bell-shape such a Gaussian or cosine, but this advantage is outweighed by the double counting problem.

A suitable 2nd derivative filter can be obtained by subtracting the mean from any smoothing filter. Savtizky-Golay filter provide a family of 2nd derivative filters. The 2nd order filter is most suitable.

The preferred embodiment is an apodized Savitzky-Golay filter (ASG), which has very smooth tails in the transfer function. The filters I use are a cosine smoothing filter, and an cosine-apodized 2nd order polynomial Savitzky-Golay 2nd derivative filter. The smooth tails reduce double counting due to noise.

Preferred characteristics of convolution filters

The elements of convolution of this convolution matrix C are chosen to correspond to the typical shape and width of the ion. For example, the cross section of the central row of C matches the chromatographic peak shape; the cross section of the central column of C matches the spectral peak shape.

Disadvantages of the Gaussian Matched Filter

The filter of the preferred method is not the GMF. To motivate the choice of the filter that is preferred, we first specify three disadvantages of the GMF.

First, the GMF will produce a widened output peak for each ion. Second, a Gaussian filter has only positive coefficients, and thus will preserve the baseline response underlying each ion. Finally, such a MGF requires a large number of multiplications to compute each data point in the output matrix.

The preferred filter for the method is not the GMF prescribed by the MFT. But the preferred filter does satisfy the requirement that it produce only a single local maximum for each ion peak and does address these limitations of the GMF.

Regarding peak broadening, it is well know that if a signal, which has positive values and has standard width σ_s, is convolved with a filter, which has positive values and whose standard width is σ_f, then the standard with of the convolved output is increased. The signal and filter width combine in quadrature to produce an output width of $\sigma_o = \sqrt{\sigma_s^2 + \sigma_f^2}$. In the case of the GMF, where the widths of the signal and filter are equal, the result is for the output peak to be a factor of $\sqrt{2} \approx 1.4$, or 40% more broad then the input peak.

Peak broadening will case the apex of a small peak to be masked by a large peak, when the small peak is nearly coeluted in time and nearly coincident in mass-to-charge with the larger peak. A simple way to address this issue is to simply reduce the width of the Gaussian convolution function. For example, halving the width produces an output peak that is only 12% more broad. The peak widths will no longer be matched, the SNR will be reduced, but more of the nearly coincident peak-pairs will be detected.

Regarding baseline preservation, a positive-coefficient filter will always produce a peak whose apex amplitude is the sum of the actual peak amplitude plus the underlying baseline response. Such a background, baseline intensity can be due to a combination of detector noise and other low-level peaks, sometimes termed chemical noise. In order to obtain a accurate measure of amplitude, a baseline subtraction operation must be employed. Such an operation would require a separate algorithm to detect the baseline responses surrounding the peak, interpolate those responses to the peak center, and subtract that response.

The method that we propose here can accomplish baseline subtract by considering filters that have both negative, as well as positive coefficients. These filters can be termed deconvolution filters, and are implemented by filter coefficients that are similar in shape to filters that extract the second derivatives of data. We will show examples of such filters that will produce a single local-maximum response for each ion. Another advantage of such filters is that provide a measure of deconvolution, or resolution enhancement. Not only will they preserve the apex of peaks that appear in the original data matrix, but they will also produce apices for peaks that are were visible only as shoulders, not apices, in the original data.

Regarding the computational burden, the straightforward convolution that uses a GMF is inefficient and slow as compared to other filter formulations. For example, if we choose a 20 point wide filter in the time direction and a 20 point wide filter in the spectral direction, then each output point requires $20 \times 20 = 400$ multiplications and additions.

In the following sections, we will address each of these points in turn. We will introduce two styles of convolution filters that are computationally more efficient and a 2^{nd} derivative deconvolution filter. We will conclude by describing the preferred filter for the method, which we will term a rank-2, combined smoothing and 2^{nd} derivative filter.

Rank-1 Convolution Filters

Until now, the convolution filters we described were matrices that contained $P \times Q$ independently specified coefficients. We now describe two other ways that these filter coefficients can be specified. The resulting convolution coefficients are not as freely specified, but the computation burden is eased. This section describes a rank-1 implementation of a two-dimensional convolution.

A two-dimensional convolution of the LC/MS data matrix can be accomplished by the successive application of two one-dimensional convolutions. Consider a one-dimensional filter g_q that is first applied to each column of the LC/MS data matrix, producing an intermediate convolved matrix. To this intermediate matrix, we apply a second one-

dimensional filter f_p, this time to each row. Each one dimensional filter can have a different set of filter coefficients. The first expression shows how these filters can be applied in succession, where the intermediate matrix is enclosed in the braces.

$$c_{i,j} = \sum_{p=-h}^{h} f_p \left(\sum_{q=-l}^{l} g_q d_{i-p,j-q} \right)$$
$$= \sum_{p=-h}^{h} \sum_{q=-l}^{l} f_p g_q d_{i-p,j-q}$$

The first expression specifies how the method is implemented. If f_p contains P coefficients and g_q contains Q coefficients, then the number of multiplications needed to compute a value for $c_{i,j}$ is $P+Q$. Thus in the case where $P=20$ and $Q=20$, then only 40 multiplications are needed for each point. This is in contrast to the general case where $20\times 20 = 400$ are needed for each $c_{i,j}$.

The second expression is a rearrangement that shows that the successive operations are equivalent to a convolution of the data matrix a single coefficient matrix whose elements are pair-wise products of the one dimensional filters. It is this second expression that shows why we term this a rank-1 formulation. This is because the effective two-dimensional convolution matrix is a rank-1 matrix formed by the outer product of two one-dimensional vectors. Thus we can write the second expression as follows

$$c_{i,j} = \sum_{p=-h}^{h} \sum_{q=-l}^{l} F_{pq} d_{i-p,j-q}$$
$$F_{pq} \equiv f_p g_q$$

and it is matrix F_{pq} that two-dimensional coefficient matrix that will emerge from the convolution operation.

The rank-1 filter is described by two orthogonal cross sections, one for each filter. The filter for each orthogonal cross-section is specified by a one-dimensional filter array.

As an example of the application of this formulation, we can choose f_p and g_q to have Gaussian profiles. The resulting F_{pq} will have a Gaussian profile in each row and column. The values for F_{pq} will be close, but not identical to $f_{p,q}$ for the GMF. Thus this rank-1 formulation can give results similar to the GMF, but with a reduction in computation time by a factor of $400/40 = 10$ in our example.

Smoothing and Second derivative Filters

The coefficients of the convolution matrix can be chosen to perform smoothing operation or deconvolution operations, or some combination. The smoothing characteristics of the convolution matrix address the problems produced by system noise. The deconvolution characteristic address the problems produced coelution and interference.

The one-dimensional Gaussian filter is an example of a smoothing filter. A general characteristic of smoothing filters is that their coefficients sum to a positive non-zero number. Conventionally, the coefficients are normalized so that their sum equals unity.

Other examples of filters that will smooth data are boxcar filters that have N-constant coefficients. Other examples of smoothing filter are those that have triangular, trapezoidal, parabolic, or co sinusoidal cross-sections.

The parabolic smoothing filter is a special example of a family of filters that are specified by sums of weighted polynomial shapes. The commonly cited reference to these filters is the article by Savitzky-Golay. Savitzky-Golay (SG) filters can describe a family of one-dimensional filters that can smooth data.

Other important categories of ond-dimensional filters are those that differentiate data. Though such filters can be assembled from combinations of box, triangle, and trapezoidal shapes, the most common specification of filters that differentiate data are SG polynomial filters.

Below we describe a modified version of SG smoothing and differentiating filters, called Apodized Savitzky-Golay filters (ASG). Every SG filter has a corresponding ASG filter. The ASG filters provide the same basic filter function as the SG filter, but an ASG filter produces higher attenuation of unwanted high-frequency noise components than do corresponding SG filters.

Filters that extract the second derivative of a signal are of particular importance to ion detection. The second derivative of a signal measures its curvature. The most prominent characteristic of a peak, in one- or two-dimensions is its apex, and the apex of a peak is the point on the peak that has the highest curvature. Peaks that are shouldered represent regions of high curvature, and the second derivative of a shouldered peak can be used to detect the presence of such a peak against the background of a larger, interfering peak.

A general characteristic of the coefficients of a differentiating filter is that its coefficients sum to zero. A general characteristic of the coefficients of a second derivative filter filter is that the first moment of its coefficients sum to zero, guaranteeing that the response to a constant or straight line (having zero curvature) is zero.

Figure 15 gives examples of second derivative filters.

[Figure 15 second derivative filters.]

In the one-dimensional case, the advantage of a second derivative filter over a smoothing filter is that the amplitude of the second derivative at the apex is proportional to the amplitude of the underlying peak. The second derivative of the peak does not respond to the baseline, and thus, in effect, performs the operation of baseline subtract and correction automatically. Second derivative filters can have the unwanted effect of increasing noise relative to the peak apex. This can be addressed by presmoothing the data. The preferred method is to increase the width of the filter which in effect increases its ability to smooth.

Rank-1 Smoothing and differentiating filter

The rank-1 filter allows us to choose a separate filter for each dimension. Here we consider a pair of filters that address the problems associated with the GMF. We choose g_q to be a second derivative filter, and we choose f_p to be a smoothing filter.

We choose the f_p smoothing filter (applied to the spectral direction) to be a cosinusoidal filter, whose fwhm is about 70% of the fwhm of the corresponding mass peak. We choose the g_q second derivative filter (applied to the spectral direction) to be an ASG filter, whose zero crossing width is about 70% of the fwhm of the corresponding chromatographic peak.

The cross-sections of these filters are in Figure 15.

These filters are then applied to the data according to Equation xx.

The advantage of this formulation over the GMF is that is fast, because it is a rank-1 filter; that it provides a linear, baseline corrected response that can be used for quantitative work; and that is sharpens, or partially deconvolves fused peaks in the chromatographic direction.

The filter functions of f_p and g_q can be reversed. We can choose to make f_p the second derivative filter and make g_q the smoothing filter. Such a rank-1 filter would then deconvolve shouldered peaks in the spectral direction, and smooth in the chromatographic direction.

Note that it is not possible to choose both f_p and g_q to be second derivative filters. The rank-1 product matrix that gives the equivalent convolution matrix contains not one, but four local maxima when convolved with an ion peak. The four additional positive apices are side-lobes that come from the products of the negative lobes associated with these filters. Thus this particular rank-1 filter will not be suitable for the proposed method.

In summary, for the method proposed here, a rank-1 convolution operation can be carried out by a filter that a either a smoothing or 2-d cross section, in the following combinations

m/z	**Time**
Smoothing	Smoothing
Smoothing	2nd derive
2nd derive	Smoothing

Rank-2 Convolution Filters

It is desirable to have a convolution filter whose cross-section is that of a second derivative filter in both the chromatographic and spectral directions. This could be accomplished with a modified version of the GMF that could be formulated as follows:

$$f_{i,j} = \exp\left(-\frac{1}{2}\frac{(i-i_o)^2}{\sigma_c^2}\right)\exp\left(-\frac{1}{2}\frac{(j-j_o)^2}{\sigma_c^2}\right)$$

This filter addresses the baseline correction problem and the deconvolution problems associated with the GMF. But the issue with this two-dimensional filter is its computational burden. All coefficients need to be multiplied to determine each output value.

We introduce the rank-2 convolution filter to do that. The rank-2 filter is obtained by computing two rank-1 filters and summing their result. Thus there are four filters; two associated with the first rank-1 filter and two associated with the second rank-1 filter. These four filters f_p^1, f_r^2 and q_q^1, q_s^2 are used to form the ouput as follows:

$$\begin{aligned} c_{i,j} &= \sum_{p=-h}^{h} f_p^1\left(\sum_{q=-l}^{l} g_q^2 d_{i-p,j-q}\right) + \sum_{p=-h}^{h} f_p^2\left(\sum_{q=-l}^{l} g_q^2 d_{i-p,j-q}\right) \\ &= \sum_{p=-h}^{h}\sum_{q=-l}^{l}\left(f_p^1 g_q^1 + f_p^2 g_q^2\right) d_{i-p,j-q} \end{aligned}$$

The first expression shows how each filter pair can be applied in succession, where the intermediate matrix is enclosed in the braces, and how the results from the two rank-1 filters are summed.

The second expression is a rearrangement that shows that the successive operations are equivalent to a convolution of the data matrix a single coefficient matrix whose elements are sum of pair-wise products of the two one-dimensional filter pairs.

The first expression specifies how the method is implemented. If f_p^1 and f_p^2 both contain P coefficients and g_q^1 and g_q^2 both contain Q coefficients, then the number of multiplications needed to compute a value for $c_{i,j}$ is $2(P+Q)$. Thus in the case where $P = 20$ and $Q = 20$, then only 80 multiplications are needed for each point. This is in contrast to the general case where $20 \times 20 = 400$ are needed for each $c_{i,j}$ for the more general case.

We term this a rank-2 formulation because the effective two-dimensional convolution matrix is a rank-2 matrix formed by the sum of the outer product of two pairs of one-dimensional vectors. Thus we can write the above expression as follows

$$c_{i,j} = \sum_{p=-h}^{h}\sum_{q=-l}^{l} F_{pq} d_{i-p,j-q}$$

$$F_{pq} \equiv f_p^1 g_q^1 + f_p^2 g_q^2$$

and it is the two-dimensional coefficient matrix F_{pq} that will emerge from the convolution operation.

Rank-2 Smoothing and differentiating filter

The rank-2 filter allows us to choose two filters for each of two dimensions. Here we specify four filters that address the all problems associated with the GMF. These problems are baseline correction, deconvolution in both chromatographic and spectral directions, and computational efficiency. The filter described now is the preferred filter for the ion detection and quantitation method.

We choose f_p^1 to be a smoothing filter (applied to the spectral direction) implemented as a cosinusoidal filter, whose fwhm is about 70% of the fwhm of the corresponding mass peak. We choose g_q^1 to be a second derivative filter (applied to the spectral direction) to be an ASG second-derivative filter, whose zero crossing width is about 70% of the fwhm of the corresponding chromatographic peak.

We choose f_p^2 to be a 2nd derivative filter (applied to the spectral direction) implemented as a 2nd derivative ASG filter, whose zero-crossing width is about 70% of the fwhm of the corresponding mass peak. We choose g_q^2 to be a smoothing filter (applied to the spectral direction) implemented as a cosinusoidal filter, whose fwhm is about 70% of the fwhm of the corresponding chromatographic peak.

The cross-sections of these filters are in Figure 15.

These filters are then applied to the data according to Equation xx.

The advantage of this formulation over the GMF is that is fast, because it is a rank-2 filter; that it provides a linear, baseline corrected response that can be used for quantitative work, because each cross-section is a second-derivative that sums to zero; and that is sharpens, or partially deconvolves fused peaks in the chromatographic direction, again, because each cross-section is a second derivative filter.

The properties and uses of the ion parameters and the ion table

The output of the detection and quantitation method is a list of ions and their properties. This list can be described as a table with least three and possibly more columns. The first three entries in each row in the table are an ions retention time, mass-to-charge ratio, and intensity. Other rows may contain, for example, the ions width as measured by the fwhm of the convolved peak, or the zero-crossing from the convolved peak. The smoothing filter measures the fwhm of the peak. The second derivative filter measures the zero-crossing width. The number of rows in the table corresponds to the number of ions detected.

Storage method

The computer memory needed to store the information contained in the ion table is much less than the memory needed to store original LC/MS data. A typical injection that contains 3600 spectra (collected once per second for an hour), with 400,000 resolution elements in each spectrum (20,000:1 MS resolution, from 50 to 2,000 amu) will require in excess of several giga-bytes to store the LC/MS data matrix of intensities.

For complex sample, the ion detection and quantitation method proposed here can detect upwards of 100,000 ions. These ions are represented by a table that 100,000 entries, with its distinct retention time and m/z and intensity triplet. The amount of computer storage to represent such a table is less than 100 megabytes, several percent of the memory needed to store the original data. Thus less storage is required of the data in the tabular form than compared to the original LC/MS data matrix. Thus the ion detection and quantitation method that leads to the ions table also leads to a storage method. The storage method consists of saving the information in the ion table in a form that can then be accesses and extracted for further processing.

Methods for the analysis of complex LC/MS data

Separation of a mixture by LC simplifies the spectra obtained by the MS analysis. Consider a mixture that contains, say 1000, molecular species that may produce detectable ions. If the mixture were injected into the MS all at once, the spectra may contain 1000 ions, some of which may overlap and interfere. If the MS analysis is preceded by an LC separation, then the number of ions seen in a given spectra can be greatly reduced. For example, at any one moment, there may be only, say 10 molecules present in the flow-cell. Thus the spectra obtained for that moment would have at most 10 ions present.

By detecting ions according to the method described about, we can further reduce the complexity of the spectra. Even though there may be only 10 molecules in the flow cell, each may be eluting at different retention times. For example, three of these molecules are from peaks that are just starting to elute from the column. Four may be in the flow-cell in the middle of their elution profile. The final three molecules may be from the tail of peaks that are exiting the flow cell.

Biological spectra

Biological samples are one important example of mixtures analyzed by LC/MS that contain complex molecules.

A singular molecular species may produce several ions. Peptides occur naturally at different isotopic states, so a peptides that appears at a given charge will appear at several values of m/z. With sufficient resolution, the mass spectrum of a peptide shows a characteristic ion cluster.

Proteins, which have high mass, will be ionized into different charge states. The isotopic variation in proteins may not be resolved by a mass spectrometer. But the ions that appear in different charge states can be resolved, again producing a characteristic pattern.

Mass spectrometers measure only the ratio of mass-to-charge, not mass by itself.

It is possible to infer the charge state, and hence mass of molecules, such as peptides and protein from the pattern of ions they produce. For example, if a protein occurs at multiple charge states, then it is possible from the spacing of m/z values to infer the charge, and hence mass of each ion, and the mass of the uncharged parent. Similarly, for peptides, where the m/z changes due to change in the isotopic value for m, it is possible to infer the charge from the spacing between adjoining ions.

There are a number of techniques in the prior art that utilize the m/z values from ions to infer the charge and parent mass. Common to all these methods is the need to select the correct ions and to use accurate values for m/z.

The ions from the list provide high precision values. Thus these values, when employed by the methods cited, will produce results with enhanced precision.

Moreover, several of the cited methods attempt to deal with the complexity of spectra by citing various techniques to distinguish between ions that may appear in a spectrum.

The method described her, focusing in on a restricted retention time region, reduces complexity of spectra. This method described here will not remove all unrelated ions. But by eliminating many, it will provide a more simplified and easier to interpret spectrum to the methods of the prior art.

IN both these cases, the ions will appear in the LC/MS data matrix at exactly the same retention time. In the case of the peptides, peptides that differ only by isotopic mass will interact with the column identically and elute at a common retention time. Peptides and proteins that are ionized into different charge states will have eluted at the same time.

It is very useful to be able to examine the ions from an LC/MS chromatogram and identify those ions that share a common retention time.

In the methods of the prior art this is accomplished by simply selecting a spectrum centered on a prominent peak (or combining spectra associated with a peak), to obtain a single extracted MS spectra. If that peak were from a molecule that produced multiple, time-coincident ions, then the spectra would contain all those ions.

However, that spectrum will also contain ions from unrelated species. These would be species that, coincidently, elute at the exact same retention time as the species of interest.

But more frequently, these ions may be from species that elute at different retention times. But if these retention times are within about a FWHM of the chromatographic peak width, the ions from the front or tails of these peaks will appear in the spectrum.

The appearance of these peaks requires subsequent processing to recognize and remove them. At worst, they may coincide, thereby biasing measurements.

Figure 16 shows the LC/MS data matrix that results from two parent molecules, and the resulting multiplicity of ions. These two species overlap chromatographically. The figure also shows the resulting complex spectrum. These spectra were obtained by simply extracting the column spectra from the LC/MS data matrix.

These spectra could also have been obtained from coadding or combining spectra from the LC/MS data matrix.

Method for simplification of complex spectra

Here we disclose a novel method to obtain spectra from an LC/MS chromatogram. We have already described how we form the ions list by the operations of convolution, apex detection, parameter estimation, and threshold rejection. We can now select ions from a narrow retention time window. The width of this window will be no larger than the FWHM of the chromatographic peak. In practice, it can be 10 times smaller. Thus for example, we may start with the most intense ion in the list, note its retention time, and then select all ions that are within a narrow window from this retention time.

This window is chosen to include all ions that elute nearly simultaneously, and are thus candidates for being related. As importantly, this window exclude molecules that cannot be related. Many of these molecules would have been included in spectra obtained from the prior art. Thus the results spectra obtained from the peak list contains all the ions of interest, and produces a significantly simplified spectrum by excluding ions that cannot be related, by virtue of their discrepant retention time. Figure 16 shows the resulting simplified spectra that result from this targeted extraction.

[Figure 16. Example of two coeluting parent ions that each produce multiple ions]

Method for simplification of complex chromatograms

Application to chromatograms employing alternating configurations

As a sample is collected with an LC/MS system, it is important that some number of spectra be collected across the peak in order for the retention time to be accurately inferred.

For example, 5 spectra per FWHM is adequate to the task. If the LC/MS systems permits a higher sample rate, then the retention times of peaks can, of course still be adequately inferred.

It is possible to alternate the configuration of an LC/MS system on a spectrum by spectrum basis. For example, all even spectra can be collected as described above. All odd, interleaving spectra could be obtained with the MS in a different mode. One

example of a mode change is provided by LC/MS/MS where fragmentation could be performed. Thus spectra that are collected are of the fragments of the ions collected in the un-fragmented state.

Note that the fragment, or modified ions would appear with a chromatographic profile having the same retention time as the unmodified ions. (The extra time to perform modifications in online MS is short as compared to the peak width or FWHM of a chromatographic peak. The transit time of a molecule in an MS is milli- or micro-seconds. The width of a chromatographic peak is seconds or minutes.)

The unmodified and modified spectra can then be segregated into two independent data matrices. The operations of convolution, apex detection, parameter estimation and thresholding can be applied independently to both.

The result of the analysis will be two lists of ions. However, ions appearing in the list will bear some relationship. For example, an intense ion in the unmodified list may have counterpart in the list of modified ions. In that case these ions will share a common retention time. Again, a window restricting the retention time would be applied to both data matrices.

Figure 17 illustrates this point. The lower plot shows three ions detected by spectra that are unmodified. The upper plot shows eight ions that arose as a result of modifications to the MS. Ions in the upper chromatogram that are related to those in the lower appear at the same retention time, as indicated by the three vertical lines labeled to, t1, and t2.

[FIGURE 17 Fragmentation peaks that occur at retention times corresponding to precursor ions.]

Clearly this process could apply to more than two chromatograms, as long as the retention time of peaks can be estimated from the data.

The ion list

Because it are only these values associated with each apex that are of interest, the convolved matrix is then further simplified into a simple, tabular list. Each row in the table corresponds to the apex properties of a single ion. For example, we may place in the first column the m/z value, and place the retention time in the second column and the intensity (apex height) in the third column. Other columns can include other parameters of an apex that can be obtained from the convolved matrix, which, for example, can include characteristics of the shape of the convolved peak.

This list can be interrogated to form novel and useful spectra. For example, selection of ions from the table based upon the enhanced estimates of retention times produces spectra of greatly reduced complexity. These spectra because it focuses on a restricted region of retention time can exclude ions unrelated to the species in question.

Selection of ions based upon the enhanced estimates of mass-to-charge ratios produces chromatograms of greatly reduced complexity.

Retention-time selected spectra can simply the interpretation of mass spectra of molecular species that induce multiple ions in a spectrum. Examples of molecular species

that produce multiple ions at a common retention time are proteins, peptides, or their fragmentation products.

Method for Peak purity

When a chromatographic separation is followed by a single channel of detection, it is impossible to determine if a peak contains a coelutant. With an LC/MS separation, the ability to obtain spectra for each peak allows the detection of molecules that ionize. The ion list provided by the method described here allows immediate determination of ions that elute within any given time range.

Thus if the analyst wishes to determine how many compounds or ions elute within the time of a principle peak of interest, one can interrogate the ion list. For example one can come up with a figure of merit that describes the purity of a peak as follows:

Sum ions within a time range. Ratio of major ion to sum. 100% is pure, less than 100% is not pure. Use threshold to determine level of significance.

APPLICATION FOR UNITED STATES LETTERS PATENT

by

MARC V. GORENSTEIN

ROBERT S. PLUMB

and

CHRIS L. STUMPF

for a

APPARATUS AND METHOD FOR IDENTIFYING PEAKS IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY DATA AND FOR FORMING SPECTRA AND CHROMATOGRAMS

SHAW PITTMAN LLP
1650 Tysons Boulevard
McLean, VA 22102-4859
(703) 770-7900
Attorney Docket No.: WAA-347

APPARATUS AND METHOD FOR IDENTIFYING PEAKS IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY DATA AND FOR FORMING SPECTRA AND CHROMATOGRAMS

BACKGROUND

Field of the Invention

[0001] The present invention relates generally to liquid chromatography and mass Spectrometry. More particularly, the present invention relates to detection and quantification of ions from data collected by an LC/MS System.

Background of the Invention

[0002] Mass spectrometers (MS) are widely used to identify and quantify molecular species in a sample. When a sample is introduced into the MS, the molecules are ionized thereby forming ions, and the ions introduced into a mass analyzer. The mass analyzer measures the mass-to-charge ratio (m/z) and intensity of ions.

[0003] A mass spectrometer is limited as to the number of ions it can reliably detect and quantify within a single spectrum. As a result, a single complex sample injected into an MS may produce spectra too complex to interpret or analyze.

[0004] A common technique to reduce the complexity of such spectra is to precede the MS with a chromatographic separation. Such chromatographic separation can be carried out using a gas chromatograph (GC) or liquid chromatograph (LC), giving rise to both the GC/MS and LC/MS methods.

[0005] In an LC/MS system, the sample is injected at a particular time. The LC subsequently causes the sample to elute over time. The eluent is continuously introduced into the ionization source of the mass spectrometer. As the separation

progresses, the composition of the mass spectrum evolves over time, reflecting the changing composition of the eluent.

[0006] At regularly spaced time intervals, a computer-based system samples and records the spectrum seen at that interval on a storage device, such as a hard-disk drive. Typically, it is after the LC separation is complete, that the acquired spectra are analyzed.

[0007] Samples analyzed using LC/MS methods generally contain more than one molecular species. Biological samples, for example, may contain thousands, tens of thousands or more molecular species. Each molecular species may produce more than one ion. For example, the mass of a peptide depends on the isotopic forms of its nuclei and an electrospray interface can ionize proteins into families of charge states.

[0008] Reducing the number of ions simplifies the interpretation of spectra. For example, peptides or proteins can produce clusters of ions that elute at a common time and may overlap in spectra. The interpretation of such clusters is simplified if the clusters from the different molecules are separated in time.

[0009] In addition, the concentration of a species can vary over a wide range. In biological samples, it is often the case that there are more species by number at lower concentrations than at higher concentrations. It follows then that a significant fraction of ions will appear at low concentration, near the detection limit of the LC/MS. The problem of detecting low abundance species is simplified if few or species are present in the spectrum at any one time, and if the background noise present in the LC/MS chromatogram is as reduced as much as possible.

[0010] When analyzing spectra or chromatograms generated by conventional LC/MS systems, the goal is to locate, that is detect, peaks associated with ions. In conventional LC/MS systems, the spectra and chromatograms are one-dimensional. Mass (m/z) estimates for an ion are derived by examining a spectrum that contains that ion. Retention time estimates for an ion are derived by examining a chromatogram that contains that ion. The response (or intensity) of an ion can be obtained as the height or area of the peak as seen in either trace.

[0011] A common conventional technique for detecting ions is to form a total ion chromatogram (TIC), or subsets of a TIC. Conventionally, this technique is applied if there are relatively few ions to be detected. To create a TIC, all the responses collected over all m/z values within each spectral scan are summed. The sums are then plotted against the scan time. For example, a time range corresponding to the expected FWHM of a peak can be used over which to perform signal-averaging. For simple mixtures, each ion might appear as distinct peak in the TIC. However, due to ion co-elution, even in simple mixtures each isolated peak seen in the TIC may not be due to a unique ion. To help isolate peak, the apex of one peak, is selected from the chromatogram and the spectrum collected at that time is display. The resulting spectral plot is a series of mass peaks, presumably each corresponding to a single ion. A channel in the chromatogram (single m/z) corresponding to a particular peak of interest can be plotted to provide additional assurance that the peak corresponds to a single ion. The location of the peak apex in the single channel chromatogram provides an ion's retention time. The location of the peak apex in the single scan spectrum provides the value for m/z.

[0012] For more complex mixtures, in which most molecules tend to co-elute, spectral responses can generally be summed only over a subset of the collected channels, e.g, by restricting the range of m/z channels that are summed. The summed chromatogram provides information as to the ions that were detected within the restricted m/z range. Spectra can be obtained for each chromatographic peak apex, and chromatograms can be obtained for each spectral peak apex. To identify all ions in this manner, multiple summed chromatograms are generally required.

[0013] Another problem making peak detection more difficult is detector noise. A technique for identifying peaks where detector noise obscures peaks, is to signal-average the spectra or signal-average the chromatograms. Such signal averaging tends to mitigate the effects of noise. For example, the spectra that encompass a chromatographic peak can be co-added to reduce the effects of noise. The m/z values and areas and heights can be obtained from this averaged spectrum. Similarly, co-adding chromatograms centered on the apex of a spectral peak can produce chromatograms with less noise, providing more precision estimates of retention time and areas and heights.

[0014] To obtain the peak parameters, the retention time, m/z and intensity, generally peak-finding and parameter extraction algorithms are applied to each extracted or averaged trace. For example, given a spectrum, conventional techniques typically use a centroiding or peak detection algorithm to provide intensity and m/z estimates. The algorithm extracts the retention time and the peak height or area from the chromatographic trace. The algorithm extracts the mass-to-charge ratio and the peak height or area from the spectral trace.

[0015] Centroiding or peak detection algorithms typically take as input points lying on the up and down slopes of the respective peaks and combine these points using a fitting routine. For example, a quadratic, or parabolic fit can be applied to the top few points in each peak in order to estimate its apex location. The algorithms also typically produce four estimates of the intensity of the peak: the peak area of the spectrographic peak; the peak height of the spectrographic peak, the peak area of the chromatographic peak and the peak height of the chromatographic peak.

[0016] There are several problems with conventional peak detection algorithms that make it difficult to reliably detect ions and estimate their parameters. The detection methods are tedious if carried out manually and somewhat subjective whether carried out manually or automatically. Moreover, the most accurate value for m/z cannot be obtained from a single extracted spectrum. Similarly, the most accurate value for retention time cannot be obtained from a single extracted chromatogram.

[0017] Another problem with the simple signal-averaging schemes of conventional systems is that they do not provide estimates of retention time, m/z, or intensity having the highest statistical precision or lowest statistical variance. For example, there is no rule as to how many chromatograms to co-add or how to co-add them to provide statistically optional estimates. Co-adding too many may cause peaks; to be combined. Co-adding to a few may not reduce noise in an optimal fashion.

[0018] Moreover, conventional peak-detection techniques do not necessarily provide uniform, reproducible results for ions at low concentration, or for complex chromatograms, where co-elution and ion interference tends to be a common problem.

BRIEF SUMMARY OF THE INVENTION

[0019] Key parameters of ions, including their mass-to-charge ratio (m/z) retention time, and intensity, can be precisely and accurately estimated via convolving an LC/MS data matrix with a fast, linear, finite impulse response (FIR) filter, followed by peak apex detection and location. These key ion parameters can be further used to reduce spectral and chromatographic complexity.

[0020] Embodiments of the present invention provide complete accounting of the ions detected by an LC/MS apparatus. The chromatograms typically generated by an LC/MS apparatus contain noise, co-eluted compounds and partially resolved ions. The detection method of embodiments of the present invention reduce the effects of noise and resolve partially co-eluted compounds and unresolved ions. As a result, the present invention increases in the number of ions that are reliably detected. One form of outputting results if using embodiments if the present invention is a tabular list of key parameters associated with each detected ion. These parameters include the ion's mass-to-charge ratio, retention time, and intensity. These parameters are optimally estimated in the sense that the precision and reproducibility of these parameters is enhanced.

[0021] Using the created ion parameter table, embodiments of the present invention extract from subsets of ions that have desired properties or relationships. For example, ions from a common parent molecule typically have essentially identical retention times in an LC/MS chromatogram. The present invention facilitates identifying ions that lie within a retention time window about a parent ion thereby facilitating identification and grouping of related ions while ignoring unrelated ions.

[0022] Using the extraction method provided by embodiments of the present invention allows creation of spectra with reduced complexity. Such a reduced complexity spectrum is a significant improvement over conventional systems that simply extract a spectrum (or an average of spectra) from an LC/MS data matrix. This is because such conventionally generated spectra are usually contaminated by ions from the leading or tailing edge of peaks that are unrelated to the ions of interest. The ions retained in the windowed spectrum using the present invention can then be further analyzed by methods known in the prior art. For example these methods can be used to obtain the mass or identity of the common parent molecule.

[0023] Window thresholds can also be applied to extract ions of nearly the same mass-to-charge ratio from the table. Such extracting produces chromatograms corresponding to a desired mass-to-charge ratio which are of significantly reduced complexity compared to those generated using conventional systems. Use of the present invention provides enhanced completeness, accuracy, and reproducibility of the final experimental result by improving completeness, accuracy, and reproducibility of results obtained from a single injection. In addition, reduction in complexity further simplifies the interpretation of spectra and by spectra containing fewer ions and by reducing noise backgrounds, and by partially resolved co-eluted compounds and interfering ions.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] Figure 1 is a schematic diagram of an exemplary LC/MS system according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0025] Liquid chromatography followed by on-line mass spectrometry (LC/MS) provides a powerful means to identify and quantify molecular species in a wide variety of samples. Typical samples can contain a mixture of a few or thousands of molecular species. The molecules themselves can span a wide range of properties and characteristics.

[0026] LC/MS systems generally analyze the content of a single mixture at a time. Such analysis is commonly referred to as the analysis of an injection.

[0027] Typically a sample is only one of a set of samples to be analyzed. Tuning experiments on each sample of the sample set provides data meaningful results can be obtained. For example, a sample set can contain calibration samples, control samples, and unknown samples that are obtained under a variety of conditions. The desired result from an experiment might be a determination of how the concentration of one analyte of interest has changed between and within the controls and unknowns.

[0028] The analysis of a sample set is typically carried out by analyzing each sample in serial order. To measure the reproducibility of the results obtained from a given injection, a typical experimental protocol may require that each sample be divided and analyzed in replicate; and each sample be analyzed by different, but nominally equivalent, LC/MS systems.

[0029] LC/MS systems can analyze the content of a single mixture at time or analyze parallel separations. For example, the MUX system available from Waters Technologies, Inc., of Milford, MA allows analysis of parallel sample mixtures to

provide greater analysis throughput. The results of either class of system can be analyzed using embodiments of the present invention.

[0030] Embodiments of the present invention can be applied to a variety of applications including large-molecule, non-volatile analytes that can be dissolved in solvent. Such analytes are best separated by liquid chromatographic techniques. Although embodiments of the present invention are heretofore described only with respect to LC or LC/MS, the ion detection and analysis method disclosed herein apply to other analysis techniques, such as GC or GC/MS analysis as well.

[0031] Figure 1 is a schematic diagram of an exemplary LC/MS system 101 according to an embodiment of the present invention. An LC/MS environment or analysis is performed by injecting a sample 102 into a liquid chromatograph 104. The injection can be by manual or automatic means. A high pressure stream of chromatographic solvent forces sample 102 to migrate through a chromatographic column 106. Column 106 typically comprises a packed bed of silica beads to whose surface comprises are bonded molecules that determine the migration velocity of each molecular species. The resulting migration time of a species depends upon competitive interactions between that molecule, the solvent, and the beads.

[0032] A species migrates through column 106 and emerges, or elutes, from column 106 at a characteristic time, conventionally referred to as the molecule's retention time. Once the molecule elutes from column 106, it can be conveyed to a detector, such as a mass spectrometer 108.

[0033] A retention time is an average time. A molecule that elutes from a column at retention time t actually elutes over a period of time that is centered at time t. The

elution profile is termed a chromatographic *peak*. The elution profile of a peak is typically bell-shaped, and has a width. The peak's width is described by its full width at half height, or half-maximum (FWHM).

[0034] The peak width, as measured by FWHM, is independent of the height of the peak and is substantially a constant characteristic of a molecule for a given separation method. Ideally, for a given chromatographic method, all molecular species elute with the same peak width. In practice, peak widths change with retention time. For example, molecules that elute at the end of a separation may display peak widths that are two times wider than those associated with molecules that elute early in the separation. Thus, as measured by FWHM, peak widths can vary by a factor of 2 or 3 or more. Embodiments of the present invention accommodate the range of peak widths typically encountered in a chromatographic separations.

[0035] Although chromatographic separation is a continuous process, a detector that receives the eluent typically samples the eluent at regularly spaced intervals. The rate or interval at which a detector samples the eluent is commonly referred to as the sample frequency or period. The chromatographic peak width determines the minimum sample period because the sample period must be long enough so that the system adequately samples the profile of each peak. In an embodiment of the present invention, the sample period is set to make approximately five (5) measurements during the FWHM of a chromatographic peak.

[0036] For purposes of subsequent description, are assumed to have a Gaussian profile. For a Gaussian profile, the FWHM approximately 2.35 times the standard deviation σ of the Gaussian profile.

[0037] In addition to its width, a peak (chromatographic of spectral) has a height or area. The height and the area are of a peak measures of the response of the detector to the molecular species. Generally, the height and area of the peak are proportional to the amount or mass of the species injected into the liquid chromatograph. The term intensity refers to a measure of the detector's response to the amount of the species introduced into the LC/MS system and commonly is used to refer to either the height or area of the chromatographic peak.

[0038] In an LC/MS system, the chromatographic eluent is introduced into a mass spectrometer (MS) 108. MS 108 comprises a desolvation system 110, an ionizer 112, a mass analyzer 114, a detector 116, and a computer 118. When the sample is into MS 108, desolvation system 110 removes the solvent, and ionizing source 112 ionizes the analyte molecules. The ionized molecules are then conveyed to mass analyzer 114. Mass analyzer 114 sorts or filters the molecules by their mass-to-charge ratio. Molecules at each value for m/z are then detected with detection apparatus 116. The detector response is proportional to the intensity of ions at each mass-to-charge interval. The intensity is plotted as a function of m/z is the mass-to-charge spectrum.

[0039] As mentioned above, the elution of molecules from the chromatographic system is a continuous process, and, as in any LC separation, the detector samples the eluent at a regularly spaced time interval. In an LC/MS system, the MS collects and measures mass spectra at these regularly spaced time intervals. Commonly each spectrum is referred to as a scan, and each element of the spectrum is referred to as a channel. Each spectral scan in the series of spectra output by air LC/MS system can be described by its scan time.

[0040] The mass-to-charge spectra or scans can be recorded by computer 118 and stored in a storage medium such as hard-disk drive accessible to computer 118. Typically, a spectrum or chromatogram is recorded as an array of values by computer system 118 and is stored by computer system 118 for later display and mathematical analysis. Thus, a typical LC/MS separation analysis results in a series of mass-to-charge spectra stored on a hard disk drive or other storage system.

[0041] Mass analyzers measure the ratio of a molecule's molecular weight to its charge. A molecule of molecular weight m and charge z will appear as an ion with a mass-to-charge ratio m/z in a mass spectrum. The symbol μ is used to refer to the mass-to-charge ratio. That is, $\mu \equiv m/z$.

[0042] The specific functional elements that make up an MS system, such as MS 108, can vary from LC/MS system to LC/MS system. Embodiments of the present invention can be adopted for use with any of the wide range of components that can make up an MS system.

[0043] Ionization methods to ionize molecules that evolve from LC 104 include electron-impact (EI), electrospray (ES) , and atmospheric chemical ionization (APCI).

[0044] Mass analyzers, such as Mass and analyzer, 114 that are used to analyze ionized molecules in MS 108 include quadruple mass analyzers (Q), time-of-flight (TOF) mass analyzers, and Fourier-transform-based mass spectrometers (FTMS). Mass analyzers can be placed in tandem in a variety of configurations, including, *e.g*, quadruple time-of-flight (Q-TOF) Mass analyzers. Mass analyzers can include on-line collision modification of an already mass-analyzed molecule. For example, in triple quadrupole based massed analyzers (such as Q1-Q2-Q3 or Q1-Q2-TOF mass

analyzers), the second quadruple (Q2), impresses accelerating voltages to the ions separated by the first quadruple (Q1). These ions, collide with a gas expressly introduced into Q2, and are fragmented. Those fragments are further analyzed by the third quadruple (Q3) or by the TOF. Typically, it is the ions after Q3 that are detected and have their spectra recorded. Embodiments of the invention are applicable to spectra and chromatograms obtained from any mode of mass-analysis such as those described above.

[0045] After mass-to-charge analysis, the LC/MS apparatus detects and records the ions. The detection of ions can be performed by a current measuring electrometer, or a single ion counting multi-channel plate (MCP). To facilitate the present description, an embodiment of the present invention using an MCP is assumed. In this configuration, ion detection is represented by a specific number of counts. The present invention is not limited to use of an MCP as the ion detection means, and any ion detection means can be used without loss of generality.

[0046] After the chromatographic separation is completed and the ions are detected and recorded, the data is analyzed using a post-separation data analysis system (DAS). The DAS is generally implemented by computer software executing on a computer such as computer 118 shown in Figure 1. The DAS is configured to perform a number of tasks, including providing visual displays of the spectra and/or chromatograms as well as providing tools for performing mathematical analysis on the data. The analyses provided by the DAS include analyzing the results obtained from a single injection and/or the results obtained from a set of injections to be viewed and further analyzed. Examples of analyses applied to a sample set include

the production of calibration curves for analytes of interest, and the detection of novel compounds present in the unknowns, but not in the controls.

[0047] Figure 3 is illustrates exemplary spectra of three ions: ion 1, ion 2 and ion 3 produced by LC/MS analysis of a sample. Ion 1, ion 2 and ion 3 appear within a limited range of retention time and m/z. For the present example, it is assumed that the mass-to-charge ratios of ion 1, ion 2 and ion 3 are different, and that the molecular parents of the ions eluted at nearly, but not exactly, the same retention times.

[0048] It is further assumed that the retention times are close enough so that the elution profiles of the respective molecules overlap or co-elute, but are not exactly coincident. In this case, there is a of time when all three molecules are present in the ionizing source of the MS. Spectrum B in Figure 3 is one spectrum collected when all three ions are present as peaks. Note that each spectral peak is resolved by the MS. In this case, resolution means there is no overlap. The apex location of each of ions 1, 2 and 3 represents its m/z ratio.

[0049] Although it can be determined that each of the molecules was eluting from the column at this time, it is not possible to determine the precise retention time at which each ion eluted using only spectrum B. For example, spectrum B could have been collected from the front of a chromatographic peak, as the molecule began to elute from the column, or from the tail of the chromatographic peak, when the molecule was nearly finished eluting.

[0050] By examining successive spectra, it is possible to determine the retention time of the eluting molecules or at least the elution order. For example, consider the three successive spectra A, B, and C shown in Figure 3. Spectra A was collected at time

tA. .Spectrum B was collected at a later time tB. Spectrum C was collected at an even later time tC. The elution order of the respective molecules can be determined by examining the relative heights of the peaks as time progresses from tA to tC. A review of spectra A, B, and C reveals that as time progresses ion 2 is decreasing in intensity relative to ion 1, and that ion 3 is increasing in intensity relative to ion 1. Therefore, in the example illustrated by the spectra in Figure 3, ion 2 elutes before ion 1, and ion 3 elutes after ion 1.

[0051] This elution order can be verified by the following procedure. First, from Figure 3, the m/z value at the apex of each peak is obtained. Given these three m/z values, the DAS extracts from each spectrum the intensity obtained at that m/z and plots it versus elution time. Figure 4 plots the three resulting curves, which are the chromatograms obtained at three values of m/z for ions 1, 2, and 3. As can be seen, each chromatogram contains a single peak. A review of the chromatograms for ions 1, 2 and 3 illustrated in Figure 4 confirms that ion 2 elutes at the earliest time. The apex location in each of the chromatograms shown in Figure 4 represents the elution time for the molecule corresponding to the respective ions.

[0052] **[FIGURE 4. Chromatograms for three ions.]**

[0053] The foregoing analysis suggests that rather than regard the output of an LC/MS as a series of spectra, it is advantageous to regard the output as a matrix of intensities. The matrix is constructed by placing each spectrum collected at increasing time in a successive column of the matrix with columns reflecting increasing time. Once in matrix form, each column of the matrix represents a spectrum collected at time t, and each row represents a chromatogram collected at

fixed m/z. Any-a-row-oriented cross-section is the chromatographic separation at a particular m/z, and any column-oriented cross-section is the mass-to-charge spectrum at a particular time. Step 1: Create matrix of intensities. The matrix can be oriented such that rows represent chromatograms and columns represent spectra or vice versa. The remaining disclosure assumes column-oriented spectral data.

[0054] In matrix form, the data can be examined by means of a contour plot as shown in Figure 5. In the contour plot, each of ions 1, 2 and 3 appears as an island of intensity. The contour plot distinctly shows three ions and that the elution order is ion 2, followed by ion 1, followed by ion 3. Figure 5 also shows an important role of apex location. The locations of the apices in Figure 5 correspond to the m/z and retention for each ion. The height of the apex above the zero value floor of the contour plot measures the ion's intensity. The counts or intensities associated with a single ion are contained within an ellipsoidal region. The FWHM of this region is the column direction is the FWHM of the mass peak. The FWHM of this region in the row direction is the FWHM of the chromatographic peak.

[0055] Six lines are drawn through the contour plot. Each lines corresponds to a row or column, plotted in Figures 5.1 and 5.2. The three horizontal rows are the three chromatograms corresponding to rows that traverse the apex of the respective peaks. The three vertical lines are a series of time corresponding to spectra 3A, 3B and 3C illustrated in Figure 3 respectively, the center of which corresponds to the apex of ion 2.

[0056] Once spectra are obtained, the LC/MS system attempts to detect ions recorded during an LC/MS experiment and to obtain for each ion accurate values for its

retention time, m/z, and intensity. Two additional problems arise that can interfere with the ion detection effort.

[0057] The first effect arises due to the finite width of the peaks in both the spectral and chromatographic directions. The second effect results from noise present in the instrument.

[0058] Figure 6 shows the contour plot that arises from an ion 4 that is assumed to have an m/z value somewhat larger than that of ion 1. In addition, ion 4 is assumed to have a retention time that is also somewhat larger than the retention time of ion 1. Moreover, the apex of the ion 4 is assumed to lie within the FWHM of the apex of ion 1 in both the spectral and chromatographic directions. As a result, ion 4 is coeluted with ion 1 in the chromatographic direction and interferes with ion 1 in the spectrometric direction. Figure 7 shows the resulting spectra obtained at times A, B, and C. In all spectra shown in Figure 7, ion 4 appears as a shoulder to ion 1. Also, as is apparent from the contour plot of Figure 6 there is no distinct apex associated with ion 4.

[0059] Another factor that can inhibit ion detection is noise. Two kinds of noise are encountered. One kind of noise is thermal or shot noise inherent in all detection processes. Because this noise, also referred to as detection noise, is inherent in the system, it cannot be reduced. For example, counting detectors, such as MCPs, add shot noise and amplifiers, such as electrometers, add thermal or Johnson noise. Another kind of noise is chemical noise that can inhibit ion detection. Chemical noise arises from several sources, for example, from spurious small molecules that are inadvertently caught up in the process of separation and ionization. Another source

of chemical noise is found in, complex samples that may contain molecules whose concentrations vary over a wide dynamic range and samples may also include interfering elements whose effects are more significant at lower concentrations. Together detector and chemical noise, they combine to establish a baseline noise background against which the detection and quantitation of ions is made.

[0060] Figure 8 is an exemplary contour plot in which numerically generated noise is added to an ion peak contour plot to simulate the effects of chemical and detector noise. The resulting spectral and chromatographic cross-sections are shown in Figures 9 and 10 respectively. One effect of the noise can be seen by examining the contour plot of Figure 8. The added noise causes apices to appear throughout the plot. In addition, multiple apices can be seen to lie within the FWHM of the nominal apex locations associated with ions 1 and 2.

[0061] A primary goal of any LC/MS analysis technique to detect all ions and to determine accurate and precise measurement of their retention time, m/z, and intensity. This goal is achieved when each potentially detectable ion is in fact detected, and its primary parameters, retention time, m/z, and intensity are estimated from the data. Secondary observable parameters can also be determined. These secondary parameters include the widths of the peak in the chromatographic and spectral directions.

[0062] However, as described above, a number of factors make such ion detection difficult. Coelution of molecules and interferences produced by near-coincident values of m/z between two ions may cause ions to be missed, producing false negatives. Noise may cause artifacts to be detected, producing false positives *i.e.,*

false ions detected. Once an ion is detected, estimates of its retention time, mass-to-charge ratio, and intensity are determined. Coelution, interference, and noise can hamper these.

[0063] If the data matrix is free of noise and if none of the ions interfere, then each ion produces a unique, isolated island of intensity in the contour plot, an example of which is shown in Figure 5. Concentric contours identify each island, and the inner-most contour within an each island identifies the element having the highest intensity. That element is a local maximum of intensity, meaning that its intensity is greater than that of its immediate neighboring elements. For example, in a 3x3 section of the data matrix, the intensity of the element corresponding to the local maxima is greater than its 8 nearest neighbors. This element is known as the maximal element, or simply the local maximum.

[0064] As shown in Figure 5, each island contains a single maximal element. Because the maximal element is unique, ion detection can proceed as follows: (1) interrogate each element in the data matrix; (2) identify all elements that are local maxima of intensity, and (3) label each such local maximum as an ion. The parameters of the ion are obtained by examining the maximal element. The ion's retention time is the time of the scan containing the maximal element. The ion's m/z is the m/z for the channel containing the maximal element. The ion's intensity is the intensity of the maximal element itself.

[0065] This detection and quantification algorithm, it may not be adequate in all circumstances. In the presence of noise, for example, many local maxima are due to the noise, not ions. Consequently, noise may result in false positives. Threshold

criterion can be applied to the ion's intensity to reduce these false positives. Moreover, as shown in Figure 8, the noise might produce more than one multiple local maxima for an ion. As a result, a single ion could be multiply counted. Similarity, as shown in Figure 6 a pair of ions that co-elute in time and interfere spectrally may produce only a single local maximum, not two. Thus, an ion appearing in the data matrix with significant intensity might not be counted.

[0066] Further, this method may not be statistically optimal method. The variance in the estimates of retention time, m/z and intensity are determined by the noise properties of a single element. The method does not make use of the other elements in the island of intensities surrounding the maximal element that can be used to reduce variance in the estimate. Techniques for improving the estimates are described below.

[0067] For a single-channel of data, a conventional method for reducing the effects of noise is smoothing. Smoothing can be performed by convolving the single-channel data array with a set of fixed-value filter coefficients. For example, well-known finite impulse response (FIR) fillers can be configured. With appropriate coefficients to perform a variety of operations including smoothing and differentiation. One such FIR filter that can be used to smooth or differentiate one-dimensional arrays of data is the well-known Savitzky-Golay filter (described in more detail below).

[0068] The LC/MS data matrix used in embodiments of the present invention is a two-dimensional array, in which the dimensions are retention time and m/z. A way of processing such an array is by convolving it with a two-dimensional array of filter coefficients. Other methods for applying the filter coefficients to the data matrix can

be used in embodiments of the present invention. For example, the elements of the convolution can be chosen to correspond to Savitzky-Golay smoothing or differentiation filters, among other filter shapes.

[0069] The method described above can be enhanced through application of two-dimensional filtering operation to the LC/MS data matrix. In the enhanced method: (1) the elements of the two-dimensional convolution filter are chosen in accordance with a desired filtering operation; and (2) the LC/MS data matrix is convolved with the two-dimensional convolution filter to generate an output convolved data matrix; (3) identify all local maxima in the output convolved data matrix; and (4) determine and apply a detection threshold, retaining only those local maxima whose (filtered) intensities lie above that threshold. Each retained local maximum is identified as an ion.

[0070] The parameters of each identified ion, its retention time, m/z, and intensity are obtained from the elements of the output convolved data matrix. In an embodiment of the present invention, these parameters are determined as follows: (1) The ion's retention time is the time of the (filtered) scan containing the (filtered) maximal element (2) The ion's m/z is the m/z of the (filtered) channel containing the (filtered) maximal element; (3) The ion's intensity is the intensity of the (filtered) maximal element itself.

[0071] In another embodiment of the present invention the data is massaged to provide a more precise estimate of the parameters. For example, a two-dimensional fit can be made to the elements of the convolved data matrix that surround the maximal element of the convolved matrix that corresponds to an ion. In one

embodiment of the present invention, the fit is a parabolic fit. A parabola is used because it is a good approximation to the shape of the convolved peak near its apex. Using the parabolic fit an interpolated value is found for the ion's parameters. The interpolated value provides more accurate estimates of retention time, m/z and intensity than those obtained by reading of values of scan times and spectral channels. The parabolic fitting procedure is preferably implemented with a linear-least-square optimization. The ion's retention time is interpolated as the time of the maximum of the interpolated parabola this ion's m/z is the interpolated m/z at the maximum of the parabola. The ion's intensity is the interpolated value of intensity at the maximum of the two-dimensional parabolic fit. After each ion's parameters are determined, the parameters are stored in a tabular list or table. For example, each row in the list corresponds to the parameters for a particular ion. For example, the first column corresponds to ion retention time; the second column corresponds to ion mass-to-charge ratio, and the third column corresponds to ion intensity. This list or table can then be used by other well-known processing operations as described below.

[0072] An important consideration as implementing embodiments of the present invention is the size of the convolution matrix and the values for the filter coefficients. Once the convolution has taken place, an appropriate detection threshold must be determined and applied to optimally detect and quantify ion's.

[0073] The convolution operation of embodiments of the present invention is a more general and powerful approach than the simple signal-averaging schemes of the conventional systems. The values for the convolution coefficients can be chosen to obtain values for retention time, m/z, and intensity, with better signal-to-noise ratios

obtained from the extraction of single channels or scans. Moreover, the convolution coefficients can be chosen to produce estimates of retention time, m/z, and intensity that have the greatest precision, or least statistical variance, for a particular data set. As a result, embodiments of the present invention provide more reproducible results for ions at low concentration.

[0074] In addition, the coefficients of the convolution matrix can be chosen to resolve ions that are co-eluted and/or interfering. For example, the apices of shouldered ions can be detected using embodiments of the present invention. Such detection overcomes limitations associated with conventional techniques to analyze complex chromatograms, where coelution and ion-interference are a common problem.

[0075] The convolution operation according to embodiments of the present is linear, non-iterative and open looped. Use of the convolution operation of embodiments of the present invention can provide a statistically optimum averaging of each of the components in the LC/MS chromatogram. In an embodiment of the present invention, the convolution operation is implemented by means of a general purpose programming language using a general purpose computer such as computer 118. In an alternate embodiment of the present invention, the convolution operation is implemented in a special purpose processor known as digital-signal-processor (DSP). Typically, DSP-based embodiments provide enhanced processing speed over a general purpose computer-based implementation. One advantage of embodiments of the present invention is that identification of ions as local maxima within a convolved matrix is an automatic, objective, and rapid operation.

[0076] Figure 12 is a flowchart 1202 of a method for detecting ions and establishing their parameters according to an embodiment of the present invention. In Step 1204, an LC/MS data matrix is created. As described above, the LC/MS data matrix can be created by placing LC/MS spectra collected at successive times in successive columns of a data matrix. In Step 1206 a two-dimensional convolution filter is specified according to desired filtering characteristics, which are described in more detail below. In Step 1208, a two-dimensional convolution is performed where in the LC/MS data matrix is convolved with the two-dimensional convolution filter specified in Step 1206. The output of the convolution is the output convolved data matrix illustrated in Step 1210. At this point, the ions are considered detected. In step 1213, ion parameters are determined according to the locations and intensities of the located apices. A list or table of the ion properties is created in Step 1214. Figure 13 is a graphical flowchart 1302 illustrating determination of a detection threshold and its application to the ion parameter table further consolidate the ion parameter table. In Step 1306, a detection threshold is determined and applied to the ion parameter list accessed in Step 1304 to generate an edited ion parameter list as shown in Step 1306.

[0077] The details of embodiments of the present invention described in Figures 12 and 13 are now provided more fully by first describing one-dimensional convolution more detailed description is followed by a generalization to the two-dimensional case. In general, convolution is a linear operation that combines two input arrays to produce an output array. In the present case, one of the input arrays is a data array that can vary from experiment to experiment and the other array is a set of fixed filter

coefficients. The input data array is convolved with the filter array to obtain the output array. In the one-dimensional case, for example, the input data array can be a chromatographic trace, wherein each array element represents a successive sample time. Likewise, the input data array could be a spectrum wherein each array element represents a successive m/z channel.

[0078] In one dimension, the convolution operation is defined as follows. Given a one-dimensional, *N*-element, input array of intensities d_i and filter coefficients f_i their convolution is

$$c_i = \sum_{j=-h}^{h} f_j d_{i-j}$$

where c_i is the output, convolved array.

[0079] The filter array f_j contains *M* elements. For convenience, M is chosen to be an odd number. The index *j* varies from $j = -h,\ldots,0,\ldots h$, where *h* is defined as $h \equiv (M-1)/2$. Thus, the value of c_i corresponds to a weighted sum of the *h* elements surrounding d_i. Generally M<<N. Spectra and chromatograms are examples of one-dimensional input arrays that contain peaks. The width of the convolution filer B is set to be approximately the width of the peaks. The peaks have widths $(\approx M)$,which is much smaller than the length *N* of the input array.

[0080] The index *i* for d_i ranges from 1 to *N*. However, c_i is defined only for $i \geq h$ or $i \leq (N-h)$. The value for c_i near the array boundaries, *i.e.* when $i < h$ or $i > (N-h)$, is not defined for the summation. Such edge effects can be handled by limiting the values for c_i to be those where the summation is defined. In this case,

the summation applies only to those peaks far enough away from the array edges so that the filter f_j can be applied to all points within the neighborhood of the peak. Generally, this is not a significant limitation of embodiments of the present invention.

[0081] Although the coefficients for filter f_j are typically chosen to produce a smoothing or differentiation function, embodiments of the present invention require coefficients for f_j that perform a detection function. One such set of detection coefficients is the matched filter. The coefficients are determined using the matched filter theorem (MFT). The MFT justifies use of convolution as part of a detection method.

[0082] The MFT assumes that the data array d_i can be modeled as a sum of a signal $r_o s_i$ plus additive noise, n_i.

$$d_i = r_o s_{i-i_o} + n_i .$$

[0083] The shape of the signal is fixed and described by a set of coefficients, s_i. The scale factor r_o determines the signal amplitude. The MFT also assumes that the signal is bounded, that is, it is zero (or small enough to be ignored) outside some region. The signal is assumed to extend over M elements. For convenience, M is typically chosen to be odd and the center of the signal is located at s_o. If h is defined as $h \equiv (M-1)/2$, then $s_i = 0$ for $i < -h$ and for $i > h$. In the above expression, the center of the signal appears at $i = i_o$.

[0084] For purposes of simplifying the present description the noise elements n_i are assumed to be uncorrelated Gaussian deviates, with zero mean and a standard

deviation of σ_o. More general formulations for the MFT accommodate correlated or colored noise.

[0085] Under these assumptions, signal-to-noise ration (SNR) of each element is $r_o s_i / \sigma_o$. To determine the SNR of a weighted sum of the data that contains the signal s_i an M-element set of weights w_i, where $h \equiv (M-1)/2$, and $i = -h,\ldots,0,\ldots h$ is considered. The weights are centered to coincide with the signal. The weighted sum S is defined as:

$$S = \sum_{i=-h}^{h} w_i d_{i-i_o} = r_o \sum_{i=-h}^{h} w_i s_i + \sum_{i=-h}^{h} w_i n_{i-i_o} .$$

[0086] To compute SNRs the statistical properties of the sum S are considered. Because the mean value of the noise term in an ensemble average is zero, an ensemble of S arrays is considered. The average value of S over an ensemble of arrays, where the signal in each array is the same, but the noise is different is

$$\langle S \rangle = r_o \sum_{i=-h}^{h} w_i s_i .$$

[0087] To determine the contribution due to noise, the weights are applied to a region containing only noise. The ensemble mean of the sum is zero. The standard deviation of the weighted sum about the ensemble mean is

$$\sigma \equiv \sqrt{\left\langle \left(S - \langle S \rangle \right)^2 \right\rangle} = \sigma_o \sqrt{\sum_{i=-h}^{h} w_i^2} .$$

[0088] Finally, the SNR is determined as

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\left(\sum_{i=-h}^{h} w_i s_i \right)}{\sqrt{\sum_{i=-h}^{h} w_i^2}}$$

This result is for a general set of filter coefficients w_i.

[0089] The MFT provides the values for w_i that will maximize the SNR. If the weighting factors w_i are regarded as elements of an M dimensional vector w of unit length, *i.e.,* the weighting factors are normalized so that $\sqrt{\sum_{i=-h}^{h} w_i^2} = 1$, then the SNR is maximized when the vector w points in the same direction as the vector s. The vectors point in the same direction when respective elements are proportional to each other *i.e.,* when $w_i \propto s_i$. The MFT implies that the weighted sum has the highest signal-to-noise when the weighting function is the shape of the signal itself.

[0090] If w_i is chosen such that $w_i = s_i$, then for noise with unit standard deviation, the SNR reduces to

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\left(\sum_{i=-h}^{h} s_i^2 \right)}{\sqrt{\sum_{i=-h}^{h} s_i^2}} = \frac{r_o}{\sigma_o} \sqrt{\sum_{i=-h}^{h} s_i^2}$$

These are the signal properties of the weighted sum when the filter coefficients are centered on the signal and the noise properties when the filter is in a noise-only region.

[0091] According to the MFT, to detect a signal, the convolution operation proceeds by moving the filter coefficients along the data array and obtaining a weighted sum at each point. For example, where the filter coefficients satisfy MFT, i.*e.,* $w_i = s_i$ (the filter is matched to the signal) then in the noise-only region of the data, the amplitude of the output is dictated by the noise. As the filter overlaps the signal, the amplitude

increases, and reaches a unique maximum when the filter is aligned in time with the signal.

[0092] In the two-dimensional convolution technique employed by embodiments of the present invention, the matrix of intensities output by the LC/MS experiment is one input to the two-dimensional convolution. To obtain the output convolved matrix, the LC/MS data matrix is convolved with a matrix of filter coefficients. The output convolved matrix has substantially the same number of rows and column elements as the input LC/MS matrix. Edge valve can be set to an invalid value in embodiments if the present invention to indicate invalid filtering values at the edges of output convolved data matrix.

[0093] For simplicity in the present description, assume that the LC/MS matrix is rectangular and that the size of the matrix of filter coefficients is comparable to the size of a peak. Thus, in general, the size of the filter coefficient matrix is smaller than the size of the input data matrix or output convolved matrix.

[0094] An element of the output matrix is obtained from the input LC/MS matrix as follows: the filter matrix is centered on each input element, and then the filter elements multiply the corresponding data elements and the products are summed. The procedure is described algebraically as follows.

[0095] The one-dimensional convolution operation described above can be generalized to the case of two-dimensional data. In the two-dimensional case, the input is a data matrix $d_{i,j}$ subscripted by two indices, (i, j), wherein $i = 1,\ldots,M$ and $j = 1,\ldots,N$. The data values of the input data matrix can vary from experiment to experiment. The other input array is a set of fixed filter coefficients, $f_{p,q}$, that is also

subscripted by two indices. The filter coefficients matrix, $f_{p,q}$, is a matrix that has $P \times Q$ coefficients. Variables had l are defined as $h = (P-1)/2$ and $l = (Q-1)/2$. Thus, $p = -h,\ldots,h$, and $q = -l,\ldots,l$.

[0096] Convolving of $d_{i,j}$ with $f_{p,q}$ yields the output convolved matrix $c_{i,j}$ where,

$$c_{i,j} = \sum_{p=-h}^{h} \sum_{q=-l}^{l} f_{p,q} d_{i-p,j-q} .$$

[0097] Generally, the size of the filter is much less than the size of the data matrix, so that P<<M and Q<<N. The above equation indicates that $c_{i,j}$ is computed by centering $f_{p,q}$ on the (*i,j*)th element of $d_{i,j}$ and then using the filter coefficients $f_{p,q}$ to obtain the weighed sum of the surrounding intensities.

[0098] Thus, each element of the output matrix $c_{i,j}$, obtained by the convolution operation, corresponds to a weighted sum of the elements of $d_{i,j}$, wherein each element $d_{i,j}$ is obtained from a region centered on the $_{i,j}$th element.

[0099] Edge effects can be ignored or compensated for.

[00100] The MFT discussed above for the one-dimensional case can also be generalized to the two-dimensional case for a bounded, two-dimensional signal embedded in a two-dimensional array of data. As before, the data is assumed to be modeled as a sum of signal plus noise:

$$d_{i,j} = r_o s_{i-i_o,j-j_o} + n_{i,j}$$

wherein the signal $S_{i,j}$ is limited in extent and whose center is located at (i_o, j_o) with amplitude r_o. Each noise element $n_{i,j}$ is an independent Gaussian deviate of zero mean and standard deviation σ_o.

[00101] To determine the SNR of a weighted sum of the data that contains the signal $S_{i,j}$ consider a $P \times Q$-element set of weights $w_{i,j}$, wherein $h = (P-1)/2$ and $l = (Q-1)/2$, such that $p = -h, \ldots, h$, and $q = -l, \ldots, l$. The weights are centered to coincide with the signal. The weighted sum S is defined as

$$S = \sum_{i=-h}^{h} \sum w_i d_{i-i_o, j-j_o} = r_o \sum_{i=-h}^{h} \sum w_{i,j} s_{i,j} + \sum_{i=-h}^{h} \sum w_{i,j} n_{i-i_o, j-j_o} .$$

[00102] The average value of S over the ensemble is

$$\langle S \rangle = r_o \sum_{i=-h}^{h} \sum w_{i,j} s_{i,j} ,$$

and the standard deviation of the noise is

$$\sigma = \sigma_o \sqrt{\sum_{i=-h}^{h} \sum w_{i,j}^2} .$$

and the signal-to-noise ratio is then

$$\frac{\langle S \rangle}{\sigma} = \frac{r_o}{\sigma_o} \frac{\sum_{i=-h}^{h} \sum w_{i,j} s_{i,j}}{\sqrt{\sum_{i=-h}^{h} \sum w_{i,j}^2}}$$

[00103] As in the one-dimensional case described above, the SNR is maximized when the shape of the weighting function is proportional to the signal, that is when $w_{i,j} \propto s_{i,j}$.

[00104] Again the signal properties of the weighted sum have been described under conditions where the filter coefficients are centered on the signal, and the noise properties of the weighted sum have been described where the filter is in the noise-only region.

[00105] As before, according to the MFT, to detect a signal, the convolution operation proceeds by moving the filter coefficients through the LC/MS data matrix and obtaining a weighted sum at every point. For example, where the filter coefficients satisfy the MFT, *e.g.*, $w_i = s_i$ (the filter is matched to the signal), then in the noise-only region of the data, the amplitude of the output is dictated by the noise. As the filter overlaps the signal, the amplitude increases, and reaches a unique maximum when the filter is aligned in time with the signal.

[00106] In an embodiment of the present invention using the MFT, detection proceeds as follows: (1) choose filter coefficients to correspond to the (assumed known) shape of the underlying signal; (2) convolve the data with the chosen filter coefficient; (3) identify the highest value in the convolved data; (4) check that its SNR is above a detection threshold; (5) if the filtered response meets or exceeds the detection threshold, then the signal is detected; (6) compute the arrival time and amplitude of the signal by the time and value of the local maximum in the column direction and compute the m/z as the m/z of the local maximum in the row direction.

[00107] The presence of an ion produces a peak, with a characteristic local maximum, in the convolved intensity. The detection process described above identifies peaks that satisfy a detection threshold. In one embodiment of the present invention, the detection process identifies those peaks that exceed the detection threshold as peaks

that satisfy the detection threshold. Although, embodiments of the present invention described herein identify peaks as those peaks exceeding a detection threshold, in an alternative embodiment of the present invention, the detection process identifies those peaks that meet or exceed the detection threshold as satisfying the detection threshold.

[00108] Any local maximum *i.e.,* a peak exceeding a detection threshold in the convolved output is a candidate for being a peak corresponding to a detected ion. In the absence of detector noise, every local maximum would correspond to an ion. However, in the presence of noise, some local maxima (especially low-amplitude local maxima) are due to only to the noise and not due to a genuine peak corresponding to an ion. Consequently, it is important to select detection threshold values that make it highly unlikely that a local maximum that equals or exceeds that threshold is due to noise. Following is a description of selecting an appropriate threshold.

[00109] Each ion produces a unique apex in the matrix of convolved intensities. The locations of the unique maxima in the convolved matrix provide the information on the number and properties of the ions present in the sample. As described in the method above therefore, all the local maxima of the convolved data are identified. For one-dimensional data, a local maximum is any point whose amplitude is greater than its two nearest neighbors. For two-dimensional data, a local maximum or apex is any point whose amplitude is greater than its nearest-neighbor elements. In one embodiment of the present invention, the number of nearest neighbor elements that a local maximum or apex must be greater than is eight (8). For example in the Table 1,

the central element is a local maximum because all adjoining elements have value less than 10.

8.5	9.2	6.8
9.2	10.0	8.4
7.9	8.5	7.2

Table 1: Example showing maximum

[00110] According to embodiments of the present invention, a local maximum is an ion only if the value of the local maximum exceeds a detection threshold. The value of the detection threshold can be obtained by subjective or objective means. Regardless of how the value of the detection threshold is determined, the effect of the detection threshold to divide the distribution of true peaks into two classes: those that are above the threshold and those that are below the threshold. Any true peaks below the threshold are missed by the method. Such missed true peaks are referred to as false negatives. The threshold also divides the distribution of noise peaks into two classes, those which are above the threshold and those below the threshold. Any noise peaks above the threshold are deemed ions. Such noise peaks that are deemed ions are referred to as false positives.

[00111] In embodiments of the present invention, the detection threshold is set to achieve a desired false positive rate. That is, the detection threshold is set so that the probability that a noise peak will equal or exceed the detection threshold in a given experiment is highly unlikely. The probability that a given peak above the threshold is in fact due to noise is referred to as confidence level.

[00112] To obtain fewer false positives the detection threshold is set to a higher value. However, setting the detection threshold to a higher value to reduce the incidence of false positive ion detection also means that there will be a somewhat higher false negative rate; i.e., low-amplitude, true peaks corresponding to ions will not be detected.

[00113] A subjective method for selecting the detection threshold that can be used in embodiments of the present invention is to draw a line that is close to the maximum of the observed noise. Any local maxima falling above this threshold line are considered peaks corresponding to ions. And any local maxima falling below the threshold are considered noise. Although the subjective method for threshold detection can be used, an objective criterion is preferred.

[00114] One objective method for selecting the detection threshold according to embodiments of the present invention uses a histogram of the output convolved matrix data. Figure 13 illustrates an exemplary histogram of the output convolved data matrix. The standard deviation of the intensity data in the output convolved matrix is obtained by conventional means. A threshold is chosen based on the standard deviation. As an example, two detection thresholds are set. One corresponds to 2 standard deviations. One corresponds to 4 standard deviations.

[00115] A variation of the empirical method uses the relationship between the standard deviation σ of the convolved output noise and the standard deviation σ_o of the input noise. This relationship is given $\sigma = \sigma_o \sqrt{\sum_{i=-h}^{h} s_i^2}$ assuming that the input noise is uncorrelated Gaussian deviates. Thus, the input noise can be measured, and the

standard deviation of the output can be inferred, knowing only the values used for the filter coefficients. The threshold can then be set based upon the derived output noise standard deviation.

[00116] The goal of any of the thresholding methods whether subjective or objective is to determine a detection threshold to use to edit the ion list. All ions whose intensities are below the threshold are considered noise rejected and not included in further analysis. However, no modifications are made to the values of the retained ions.

[00117] After identifying those local maxima that are peaks corresponding to ions, parameters for each peak are estimated. In one embodiment of the present invention the parameters that are estimated are the retention time, mass-to-charge ratio, and intensity. Additional parameters that can be estimated include are the chromatographic peak width and the mass-to-charge peak width.

[00118] Because the elements of the convolved matrix represent a digital sample of data, the apex of a peak in time may not coincide exactly with a sample time and the apex of a peak in mass-to-charge may not coincide exactly with an m/z channel. In general, the actual maximum of the signal in time and mass-to-charge will be offset from the available sampled values by a fraction of the sample period or the mass-to-charge channel interval. These fractional offsets can be estimated from the values of the matrix elements surrounding the element having the local maximum corresponding to the peak.

[00119] For example; in the case of one-dimensional data, a technique to estimate the fractional offset of the true apex of a peak is to locate the element of a chromatogram

or spectrum c_i that is the maximum, and then fit a parabola to this point and the two adjoining elements. The maximum of the fitted parabola corresponds to the time amplitude of the maximum value of the convolution. The time of the maximum of the parabola is an estimate of the arrival time of the peak. Both the amplitude and the arrival time obtained from this fitting procedure are optimum estimates.

[00120] An effect of the convolution of the present invention is to combine the data in the bulk of the peak so that all the information about the signal's sample and arrival time is compressed into the local maximum. Consequently, the highest element in the convolved response contains all required information about the signal.

[00121] For two-dimensional data, a technique for estimating the fractional offset of the true apex from an element of the output convolved matrix containing a local maxima corresponding to an ion is to fit a two-dimensional parabolic shape to the values of a nine-element matrix comprising the local maxima element and its eight nearest neighbors. The value of the parabola at the maximum, and its interpolated x and y values corresponding to that maximum become the estimates of ion intensity, retention time and m/z respectively. Other fits can be used within the scope and spirit of the present invention.

[00122] The interpolated location in the row direction of the maximum of the two-dimensional parabolic fit an optimum estimate of retention time. The interpolated location in the column direction of the maximum of the two-dimensional parabolic fit gives an optimum estimate of mass-to-charge ratio. The height of the apex above baseline gives an optimum estimate (scaled by filter factors) of ion intensity or concentration.

[00123] As an example of the foregoing technique, consider the case where the signal is a single peak. The peak can be modeled as a Gaussian whose width is given by the standard deviation σ_p, where the width is measured in units of sample elements. The signal is then

$$r_i = r_o \exp\left(-\frac{1}{2}\frac{(i-i_o)^2}{\sigma_p^2}\right)$$

[00124] Assume a boundary for the filter is set to correspond to $\pm 4\sigma_p$. The signal-to-noise properties of two filters are compared for the present example. The first filter to consider is a matched filter, which as described above is the signal shape itself, centered on zero, and bounded by $\pm 4\sigma_p$. The coefficients of such a matched filter are given by:

$$f_i = \exp\left(-\frac{1}{2}\frac{i^2}{\sigma_p^2}\right), \quad \text{for } i > -4\sigma_p \text{ and } i < 4\sigma_p.$$

[00125] The second filter to consider is a running average, or boxcar filter, where the coefficients of the filter are given by:

$$f_i = \frac{1}{M} = \frac{1}{8\sigma_p + 1}.$$

[00126] The output of such a boxcar filter is the average value of the input signal over M points ($M = 8\sigma_p + 1$).

[00127] For the present example, assume further that the system samples four points per standard deviation. As a result, $\sigma_p = 4$. The filters are chosen to be 33 points wide for the present example. For a Gaussian peak of unit height, the average signal

over the peak using the boxcar filter is $0.304\, r_o$, and the standard deviation of the noise is $\sigma_o / \sqrt{33} = 0.174\, \sigma_o$. Thus, the SNR using the boxcar filter is $1.75(r_o / \sigma_o)$.

[00128] For the matched filter, the maximum signal is $7.09\, r_o$, and the noise amplitude is $2.66\, \sigma_o$ for an SNR using, the matched filter is of $2.66(r_o / \sigma_o)$. As can be seen, the matched filter provides an SNR that is over 50% higher than that provided by the boxcar filter.

[00129] Both the matched filter and the boxcar filter are linear. The convolution of either of these filters with the Gaussian peak shape produces an output that has a unique maximum value. Thus, either of these filters can be used in the convolution However, because of its higher SNR at the local maximum, the matched filter is preferred.

[00130] Knowing the standard deviation of the noise and the filter coefficients, the standard deviation of the maximum value can be obtained. Using this determined standard deviation the likelihood that the maximum is due to noise can be determined. A detection threshold based upon an acceptable rate of false positives can then be set.

[00131] Linear weighting coefficients other than those that follow the signal shape can also be used. While such coefficients may not produce the highest possible SNR, they may have other counter-balancing advantages.

[00132] The ion detection and quantitation technique of the present invention provides measurements for fundamental parameters associated with each deleted ion. These parameters include the ion's retention time, m/z, and intensity. The intensity measurement estimate is simply the response of the filter output at the local maximum. The intensity measurement does not correspond exactly to peak area or

peak height. However, the intensity measurement is in proportion to those values since the convolution operation is a linear combination of intensity measurements.

[00133] The set of filter coefficients with which the LC/MS data matrix is convolved determines the scaling of the intensity. Each set of filler coefficients gives a different intensity scaling. As long as a consistent set of filters is used to determine the intensities of standards, calibrators and sample, the resulting intensity measurements produce accurate, quantifiable results regardless of the intensity scaling (described in more detail below). For example, intensities generated by embodiments of the present invention can be used to establish concentration calibration curves. Using these curves, the concentration of analytes can be estimated.

[00134] In addition to intensity, other ion properties can be obtained from the output convolution matrix generated by embodiments of the present invention. These other properties include, the widths of the ion peaks. Each peak in the output convolved matrix has a width in both the chromatographic and the spectral directions. Conventional means of measuring these widths can be applied to the peaks corresponding to each ion detected in the convolved data matrix. This provides five parameter measurements per detected ion: retention time, mass-to-change ration (m/z), intensity, peak width in the chromatographic direction and peak width in the spectral direction. Other measurements associated with detected ions are possible using the results of embodiments of the present invention. The following section will describe two types of filters that can be applied in either the chromatographic or the spectral directions. If a smoothing filter is applied, the peak width corresponds to the FWHM in that direction. If a deconvolving second derivative filter is employed, the

appropriate measure of peak width is the width between zero-crossing points, as will be described.

[00135] Because the five measurements provided by embodiments of the present invention are estimates, they each have an associated error. These errors can be estimated in a statistical sense. Two distinct factors contribute to the measurements errors. One factor to measurement error is a systematic or calibration error. For example, if the MS m/z axis is not perfectly calibrated, then any given m/z value contains an offset. Due to the nature of the systematic error, the offset is essentially constant over the entire m/z range. Such an error is independent of the signal-to-noise or amplitude of a particular ion. Similarly, in the case of m/z, the error is independent of the m/z peak width.

[00136] The second factor contributing to measurement error is the irreducible statistical error associated with each measurement. This error arises due to thermal or shot-noise related effects. The magnitude or variance of this error for a given ion depends on the ion's peak width and intensity. Statistical errors measure reproducibility and therefore are independent of calibration error. Another term for the statistical error is precision.

[00137] The statistical error associated with each measurement can in principle be estimated from the fundamental operating parameters of the instrument on which the measurement is made. For example in a mass spectrometer, these operating parameters typically include the ionization and transfer efficiency of the instrument coupled with the efficiency of the micro-channel counting plate (MCP). Together, these operating parameters determine the counts associated with an ion. The counts

determine the statistical error associated with any measurement using the mass spectrometer. For example, the statistical error associated with the five available measurements discussed above, typically follows a Poisson distribution. A numerical value for each error can be derived from counting statistics via the theory of error propagation.

[00138] In general, statistical errors can be inferred directly from the data. One way to infer statistical errors directly from the data is to investigate the reproducibility of the measurements. For example, replicate injections of the same mixture can establish the statistical reproducibility of m/z values for the same molecules. In the case of errors associated with retention time measurements, statistical reproducibility is more difficult to accomplish. The difficulty is due to systematic errors arising from replicate injections that generally mask the statistical error.

[00139] A technique to overcome this difficulty is to examine ions at different values for m/z that were produced from a common parent molecule. Since these ions originate from a common molecule in an LC/MS system the intrinsic retention time of each such ion should be identical. As a result, any difference between measurements of the retention times of such molecules must be due to statistical errors associated with the fundamental detector noise associated with measurements of peak properties. Thus, measurements of the retention time differences between ions that come from the same molecule within an injection can be used to estimate the statistical error associated with an ion's retention time.

[00140] Each measurement using an embodiment of the present invention can be accompanied by measures of its associated statistical and systematic errors. Though

these errors apply to each individual ion detected, their values can be inferred generally by analyzing sets of ions. After a suitable error analysis, the errors associated with each measurement for a detected ion can be included in each row of the table corresponding to the detected ion measurement. In such an embodiment of the present invention, each row of the table has fifteen measurements associated with each ion. These measurement are the five measurements for the detected ion corresponding to the row, the statistical and systematic errors are associated with each of the five measurements.

[00141] As described above, the statistical component of measurement error, or precision, in retention time and m/z depends on the respective peak widths and intensities. For a peak that has a high SNR, the precision can be substantially less than the FWHM of the respective peak widths. For example, for a peak that has a FWHM of 20 milli-amu and high SNR, the precision can be less than 1 milli-amu. For a peak that is barely detectable above the noise, the precision can be 20 milli-amu For purposes of the present discussion of statistical error, the FWHM is considered to be the FWHM of the peak in the LC/MS chromatogram prior to convolution, not the FWHM of the convolved peak.

[00142] Analysis of the relationship between peak width, convolution filter coefficients and signal-to-noise of the peak reveals that precision is proportional to the peak width and inversely proportional to peak amplitude. The general result can be expressed as

$$\sigma_m = k\frac{w_m}{h_p}.$$

[00143] Where, σ_m is the precision of the measurement of m/z (expressed as a standard error), w_m is the width of the peak (expressed in milli-amu at the FWHM), h_p, is the intensity of the peak (expressed as a post-filtered, signal to noise ratio), and k is a dimensionless constant of order unity. The exact value for k depends on the filter method used. This expression shows that σ_m is less than w_m, the FWHM of the peak. Thus, the present invention allows estimates of m/z for a detected ion to be made with a precision that is less than FWHM of the m/z peak width as measured in the original LC/MS data.

[00144] Similar considerations apply with respect to the measurement of retention time. The precision to which retention time of a peak can be measured depends on the combination of peak width and signal intensity. If the FWHM max of the peak is 0.5 minutes, the retention time can be measured to a precision, described by a standard error, of 0.05 minutes or 3 seconds. Using the present invention, estimates of retention time for a detected ion can be made with a precision that is less than the FWHM of the retention time peak width as measure in the original LC/MS data.

[00145] As described above, embodiments of the present invention operate by convolving an input data matrix with a filter to generate an output convolved matrix. A more detailed discussion of the filters to be used in embodiments of the present invention follows. For a Gaussian peak, the Matched Filter Theorem (MFT) specifies the Matched Gaussian Filter (MGF) as the filter whose response has the highest signal-to-noise ratio as compared to any other convolution filter. An important point is that for an individual ion the convolved output is a peak with a single local

maximum. It is the numerical value and location of each local maximum that specifies the intensity and other properties of each detected ion.

[00146] Thus, the convolution filter must have an output that produces a unique maximum when convolved with an input having a unique maximum. Since in the input signal the peak in the LC/MS data matrix has a unique maximum, the convolution filter must faithfully maintain that unique positive maximum through the convolution process. For an ion that has a bell-shaped response, this condition is satisfied by any convolution filter whose cross sections are all bell-shaped, with a single positive maximum. Examples of such filters include those whose cross-sections are bell-shaped, for example inverted parabolas, triangle filters, or co-sinusoids. Specifically, any convolution filter that has the property that it has a unique, positive valued apex makes that filter a suitable candidate to be used in embodiments of the present invention. A contour plot of the filter coefficients can be used to examine the number and location of the local maxima. All row and column and diagonal cross sections through the filter must have a single, positive, local maximum. There are a large number of filter shapes that meet this condition and that can therefore be employed in embodiments of the present invention.

[00147] Another filter shape that is acceptable is a filter having a constant value (a box car filter), since its convolution with a peak will produce an output that has a single maximum.

[00148] A well-known characteristic of boxcar filters that is advantageous in embodiments of the present invention is that such a shape produces a minimum variance for a given number of filter points. However, the transfer function for such

filters also has the undesirable characteristic of passing high frequency noise. As a result, there is a risk of double counting at low amplitude (low SNR) as a result of convolution with baseline noise that produces peaks that exceed a detection threshold. An advantage of a boxcar filter is that it can be implemented with fewer multiplications than a bell-shape filter, such a Gaussian or cosine filter. However, this advantage of boxcar filters should be considered in light of the risk of double counting described above if boxcar filters are to be used. The widths of the convolution filters can be matched to the FWHM of the peak (in time and in mass-to-charge). Such matching of filter widths is not required. For example, a Gaussian filter that has widths other than the FWHM of the ion can be used.

[00149] Another suitable class of filters that satisfy the above criterion are filters that that have a single, positive local maximum, but have negative side-lobes. Examples of such filters include filters that extract second derivatives, or curvature. The coefficient values for second derivative filters sum to zero, and this characteristic allows such fillers to be used in embodiments of the present invention.

[00150] An exemplary type filter for use in embodiments of the present invention is a smoothing filter. A suitable smoothing filter is generally a symmetric, bell shaped curve, with all positive values, and a single maximum. For example, the Savitzky-Golay polynomial filter provides a family of smoothing filters. The 0th order filter is a flat top, box car filter. The 2nd order filter is a parabola that has a single, positive maximum. Smoothing filters having asymmetric, tailed curves can also be used in embodiments of the present invention. Exemplary smoothing filters are include Gaussian shapes, triangle shapes, and parabolas, all with single maxima.

[00151] A suitable 2nd derivative filter can be obtained by subtracting the mean from a smoothing filter. Another class of suitable 2nd derivative filters are known as apodized Savitksy-Golay (ASG) filters (described in more detail below).

[00152] The elements of the filter matrix F that is convolved with the input matrix D are chosen to correspond to the typical shape and width of a peak corresponding to an ion. For example, the cross section of the central row of F matches the chromatographic peak shape; the cross section of the central column of F matches the spectral peak shape.

[00153] Although the Gaussian Matched Filter (GMF) discussed above has the characteristics corresponding to the typical shape and width of the input signal and can be used in embodiments of the present invention, it may not be optimal in all cases. One disadvantage of the GMF is that it produces a widened or broadened output peak for each ion. Another disadvantage of the GMF is that as a Gaussian filter it has only positive coefficients. Consequently, the GMF preserves the baseline response underlying each ion. A third disadvantage of the GMF is that it generally requires a large number of multiplications to compute each data point in the output convolved matrix.

[00154] To help explain peak broadening, it is well known that if a signal having positive values and a standard width, σ_s is convolved with a filter, having positive values and a standard width, σ_f, the standard width of the convolved output is increased. The signal and filter width combine in quadrature to produce an output width of $\sigma_o = \sqrt{\sigma_s^2 + \sigma_f^2}$. In the case of the GMF, where the widths of the signal and

filter are equal, the result is for the output peak to be a factor of $\sqrt{2} \approx 1.4$, *i.e.*, 40% more broad then the input peak.

[00155] Peak broadening can cause the apex of a small peak to be masked by a large peak when, for example, the small peak is nearly co-eluted in time or nearly coincident in mass-to-charge with the larger peak. One way to compensate for the possibility of such co-elution is to reduce the width of the Gaussian convolution function. For example, halving the width of the Gaussian convolution function produces an output peak that is only 12% more broad than the input peak. However, because the peak widths are not matched, the SNR is reduced relative to that achieved using a GMF. The disadvantage of reduced SNR is offset by the advantage of increased ability to detect nearly coincident peak pairs.

[00156] Regarding the second disadvantage of GMFs, that of baseline preservation, a positive-coefficient filter always produces a peak whose apex amplitude is the sum of the actual peak amplitude plus the underlying baseline response. Such background baseline intensity can be due to a combination of detector noise as well as other low-level peaks, sometimes termed chemical noise. To obtain an accurate measure of amplitude, a baseline subtraction operation is typically employed. Such an operation typically requires a separate algorithm to detect the baseline responses surrounding the peak, interpolate those responses to the peak center, and subtract that response from the peak value to obtain the optimal estimate of the peak intensity.

[00157] Embodiments of the present invention accomplish the required baseline subtraction by using filters that have negative as well as positive coefficients. Such filters are sometimes referred to as deconvolution filters, and are implemented by

filter coefficients that are similar in shape to filters that extract the second derivatives of data. Such filters can be configured to produce a single local-maximum response for each detected ion. Another advantage of such filters is that they provide a measure of deconvolution, or resolution enhancement. That is, not only do such filters preserve the apex of peaks that appear in the original data matrix, but they can also produce apices for peaks that are visible only as shoulders, not as independent apices, in the original data.

[00158] Regarding the real-time computational burden of GMFs, convolution using a GMF is inefficient and slow when compared to other filter formulations. For example, implementation of a 20 point wide GMF in the time direction and a 20 point wide GMF in the spectral direction requires $20 \times 20 = 400$ multiplications and additions per output point.

[00159] The following sections of the present description discuss alternative filter designs to the GMF that can be implemented in embodiments of the present invention. For particular applications, these filters may have better performance than the GMF.

[00160] The convolution filters described thus far are all matrices that contained $P \times Q$ independently specified coefficients. There are other ways for specifying the filter coefficients. Although the resulting convolution coefficients are not as freely specified, the computation burden is eased.

[00161] One such way of specifying the filter coefficients is through the user of a rant-1 filter implementation of a two-dimensional convolution. To understand an embodiment of the present invention implementing a rank-1 convolution filter,

consider that a two-dimensional convolution of the LC/MS data matrix can be accomplished by the successive application of two one-dimensional convolutions. Consider a one-dimensional filter, g_q, that is applied to each column of the LC/MS data matrix, producing an intermediate convolved matrix. To this intermediate convolved matrix, a second one-dimensional filter, f_p, is applied to each row. Each one dimensional filter can have a different set of filter coefficients. Equation (1) illustrates how the filters comprising a rank-1 convolution filter can be applied in succession, wherein the intermediate matrix is enclosed in the braces.

$$c_{i,j} = \sum_{p=-h}^{h} f_p \left(\sum_{q=-l}^{l} g_q d_{i-p,j-q} \right) \quad (1)$$

$$= \sum_{p=-h}^{h} \sum_{q=-l}^{l} f_p g_q d_{i-p,j-q} \quad (2).$$

[00162] Equation (1) also specifies how the method is implemented in an embodiment of the present invention. Examination of equation (1) indicates the computational burden for its implementation. If f_p contains P coefficients and g_q contains Q coefficients, then the number of multiplications needed to compute a value for $c_{i,j}$ is $P+Q$. For example, where $P=20$ and $Q=20$, only 40 multiplications are needed to determine each output point $c_{i,j}$ in the output convolved matrix. As can be seen, this is more computationally efficient than the general case where $20\times 20=400$ are required to determine each $c_{i,j}$.

[00163] Equation (2) is a rearrangement of equation (1) that illustrates that the successive operations are equivalent to a convolution of the data matrix with a single coefficient matrix whose elements are pair-wise products of the one dimensional

filters. An examination of equation (2) shows that in using the rank-1 formulation, the effective two-dimensional convolution matrix is a rank-1 matrix formed by the outer product of two one-dimensional vectors. Thus, equation (2) can be rewritten as

$$c_{i,j} = \sum_{p=-h}^{h} \sum_{q=-l}^{l} F_{pq} d_{i-p,j-q}$$
$$F_{pq} \equiv f_p g_q$$

It is that two-dimensional coefficient matrix F_{pq} that will emerge from the convolution operation.

[00164] In embodiments of the present invention using a rank-1 filter implementation, the rank-1 filter is characterized by two orthogonal cross sections, one for each filter. The filter for each orthogonal cross-section is specified by a one-dimensional filter array.

[00165] As an example of the application of a rank-1 formulation, f_p and g_q can have Gaussian profiles. The resulting F_{pq} has a Gaussian profile in each row and column. The values for F_{pq} will be close, but not identical to $f_{p,q}$ for the GMF. Thus, this particular rank-1 formulation will perform similarly to the GMF, but with a reduction in computation time. For example, in the example provided above, where P and Q were equal to 20, computational load by using the rank-1 filter is reduced by a factor of $400/40 = 10$.

[00166] As described above, the coefficients of the convolution matrix can be chosen to perform smoothing operation or deconvolution operations, or some combination. Smoothing characteristics of the convolution matrix can help address the problems

caused by system noise. The deconvolution characteristics of the convolution matrix address the problems produced co-elution and interference.

[00167] A one-dimensional Gaussian filter is an example of a smoothing filter. A general characteristic of smoothing filters is that their coefficients sum to a positive non-zero number. Conventionally, the coefficients are normalized so that their sum equals unity. Another exemplary smoothing filter is the boxcar filter, as discussed above. Other examples of smoothing filters include those having triangular, trapezoidal, parabolic, or co sinusoidal cross-sections.

[00168] For example, Savitzky-Golay (SG) describes a family of parabolic smoothing filters that are specified by sums of weighted polynomial shapes. Such parabolic filters can be used as smoothing convolution filters in embodiments of the present invention.

[00169] Another class of one-dimensional filters for use in embodiments of the present invention are those that differentiate data. Though such filters can be assembled from combinations of box, triangle, and trapezoidal shapes, the most common specification of filters that differentiate data are SG polynomial filters.

[00170] Every SG filter has a corresponding Apodized Savitzky-Golay filters (ASG) filter. An ASG filter is a modified version of an SG filter that provides the same basic filter function as the corresponding SG filter, but with higher attenuation of unwanted high-frequency noise components. Savtizky-Golay filter provide a family of 2nd derivative filters. The inventors have found that a class of filters known as apodized Savitzky-Golay filters (ASGs) work well in embodiments of the present invention. ASG filters have transfer functions that are characterized by very smooth

tails. Examples of such ASG filters include cosine smoothing filters and cosine-apodized 2nd order polynomial Savitzky-Golay 2nd derivative filters. The smooth tails are advantageous because they reduce the risk of double counting due to noise described above.

[00171] Filters that extract the second derivative of a signal are of particular use in detecting ions according to embodiments of the present invention. This is because the second derivative of a signal is a measure of the signal's curvature, and the most prominent characteristic of a peak, whether considered in one or two dimensions is its apex which is the point peak is the point of the peak that has the highest curvature. Consequently, second derivative filters can be used to enhance detection of the peaks. Moreover, peaks that are shouldered are also represented by regions of high curvature. As a result, the second derivative of a shouldered peak can be used to detect the presence of such a peak against the background of a larger, interfering peak.

[00172] A characteristic of a differentiating filter is that its coefficients sum to zero. A characteristic of a second derivative filter is that the first moment of its coefficients sum to zero. This characteristic of second derivative filters causes their response to a constant or straight line (having zero curvature) to be zero. Figure 15 illustrates the cross section of an exemplary second derivative filter in both the chromatographic and spectral directions.

[00173] In a one-dimensional case, a second derivative filter is advantageous over a smoothing filter because that the amplitude of the second derivative filter at the apex is proportional to the amplitude of the underlying peak. Moreover, the second

derivative of the peak does not respond to the baseline Thus, in effect, a second derivative filter performs the operation of baseline subtraction and correction automatically. A disadvantage of second derivative filters is that they can have the undesirable effect of increasing noise relative to the peak apex. This effect of second derivative filters can be mitigated by presmoothing the data. In one embodiment of the present invention, the width of the filter is increased. Increasing the width of the filter increases its ability to smooth the data.

[00174] Using rank-1 filters as described above, separate filtering can be applied for each dimension. In an embodiment of the present invention, for example, g_q (the filter applied in the chromatographic direction) can be a second derivative filter, and f_p (the filter applied in the spectral direction) can be a smoothing filter. Combining difference filters in difference rank-1 filter implementations can be used to overcome problems associated with filtering. For example, the rank-1 filter can be configured through appropriate consignation of filters to address the aforementioned problems associated with GMFs. For example, f_p can be a cosinusoidal filter, whose FWHM is about 70% of the FWHM of the corresponding mass peak, g_q can be an ASG filter, whose zero crossing width is about 70% of the FWHM of the corresponding chromatographic peak. Other filters and combinations of filters can be used as the rank-1 filters in other embodiments of the present invention.

[00175] The cross-sections of these filters are in Figure 15.

[00176] Using the rank-1 filter for the convolution of embodiments of the present invention has a number of advantages over the GMF. Because, it is a rank-1 filter, it is computationally more efficient than the GMF and therefore it is faster. Moreover,

the specified combination of filters provides a linear, baseline corrected response that can be used for quantitative work. Furthermore, the combination of filters sharpens, or partially deconvolves fused peaks in the chromatographic direction.

[00177] The filter functions of f_p and g_q can be reversed. That is, f_p can be the second derivative filter and g_q can be the smoothing filter. Such a rank-1 filter deconvolves shouldered peaks in the spectral direction, and smoothes in the chromatographic direction.

[00178] Note that both f_p and g_q should not be second derivative filters. The rank-1 product matrix resulting where both f_p and g_q are second derivative filters contains not one, but four local maxima when convolved with an ion peak. The four additional positive apices are side-lobes that arise from the products of the negative lobes associated with these filters. Thus this particular rank-1 filter will not be suitable for the proposed method.

[00179] Several filter combinations for embodiments of the present invention that use a rank-1 convolution filters are provided in the table below. Other filters and combinations of filters can be used as the rank-1 filters in other embodiments of the present invention.

m/z	Time
Smoothing	Smoothing
Smoothing	2nd derive
2nd derive	Smoothing

[00180] Because it might be advantageous to employ a second derivation filter as the convolution filter in both the chromatographic and spectral directions, another kind of

filtering operation can be employed in embodiments of the present invention. For example, a modified version of the GMF that could be formulated as in equation (3).

$$f_{i,j} = \exp\left(-\frac{1}{2}\frac{(i-i_o)^2}{\sigma_c^2}\right)\exp\left(-\frac{1}{2}\frac{(j-j_o)^2}{\sigma_c^2}\right) \quad (3)$$

[00181] The modified GMF of equation (3) addresses the baseline correction problem and the deconvolution problems associated with the GMF. However, one problem with implementing this two-dimensional filter as specified in equation (3) is its computational burden because all coefficients need to be multiplied to determine each output value.

[00182] To reduce this computational burden, a rank-2 convolution filter can be used. A rank-2 filter is generated by computing two rank-1 filters and summing their result. As a result, to implement a rank-2 filter in the two-dimensional convolution performed in embodiments of the present invention, four filters: $f_p^1, g_q^1, f_p^2, and\ g_q^2$ are required. Two of the filters f_p^1 and g_q^1 are associated with the first rank-1 filter and two of the filters f_p^2 and g_q^2 are associated with the second rank-1 filter. These four filters f_p^1, f_p^2 and g_q^1, g_q^2 are implemented as follows:

$$c_{i,j} = \sum_{p=-h}^{h} f_p^1 \left(\sum_{q=-l}^{l} g_q^2 d_{i-p,j-q} \right) + \sum_{p=-h}^{h} f_p^2 \left(\sum_{q=-l}^{l} g_q^2 d_{i-p,j-q} \right) \quad (4)$$

$$= \sum_{p=-h}^{h} \sum_{q=-l}^{l} \left(f_p^1 g_q^1 + f_p^2 g_q^2 \right) d_{i-p,j-q} \quad (5)$$

[00183] Filters f_p^1 and f_p^2 are applied in the spectral direction and filters g_q^1 and g_q^2 Equation (4) illustrates how each filter pair can be applied in succession, where the

intermediate matrix is enclosed in the braces, and how the results from the two rank-1 filters are summed.

[00184] Equation (5) is a rearrangement of equation (4) to show that the successive operations are equivalent to a convolution of the data matrix with a single coefficient matrix whose elements are sum of pair-wise products of the two one-dimensional filter pairs.

[00185] Equation (4) shows the preferred manner of implementing the rank-2 filter according to embodiments of the present invention. To see how implementation of a rank-2 filter eases the computational burden the filter specified is equation (3), consider that if f_p^1 and f_p^2 both contain P coefficients and g_q^1 and g_q^2 both contain Q coefficients, then the number of multiplications needed to compute a value for an element of the output convolution matrix $c_{i,j}$ is $2(P+Q)$. Thus, in the case where $P=20$ and $Q=20$, only 80 multiplications are needed to compute each element of the output convolution matrix, whereas in the general case as shown in equation (3), $20\times 20=400$ are required to compute each $c_{i,j}$.

[00186] Thus, in the rank-2 formulation, the effective two-dimensional convolution matrix is formed by the sum of the outer product of two pairs of one-dimensional vectors. Equation (4) cam be rewritten as

$$c_{i,j}=\sum_{p=-h}^{h}\sum_{q=-l}^{l}F_{pq}d_{i-p,j-q}$$
$$F_{pq}\equiv f_p^1 g_q^1+f_p^2 g_q^2$$

Two-dimensional coefficient matrix F_{pq} emerges from the convolution operation.

[00187] The rank-2 filter requires specification of two filters for each of two dimensions. In a preferred embodiment of the present invention, the four filters are specified to address the problems associated with the GMF as described above in a computationally efficient manner.

[00188] For example, in an embodiment of the present invention, the first rank-1 filter comprises f_p^1 as a smoothing filter and a second derivative filer as g_q^1 . An exemplary such smoothing filter is a cosinusoidal filter, whose FWHM is about 70% of the FWHM of the corresponding mass peak. An exemplary such second-derivative filter is ASG second-derivative filter, whose zero crossing width is about 70% of the FWHM of the corresponding chromatographic peak. The second rank-1 filter comprises f_p^2 and a smoothing filter as g_q^2 is a second derivative filter. An exemplary such second derivative filter is a second derivative ASG filter, whose zero-crossing width is about 70% of the FWHM of the corresponding mass peak. An exemplary such smoothing filter is a cosinusoidal filter, whose FWHM is about 70% of the FWHM of the corresponding chromatographic peak. Other filters and filter combinations can be used in embodiments of the present invention.

[00189] The cross-sections of these filters are in Figure 15.

[00190] The rank-2 filter described above has several advantages over the GMF. Because it is a rank-2 filter, it is more computationally efficient then the GMF and consequently faster in execution. Moreover, because each cross-section is a second derivative filter whose coefficients sum to zero, it provides a linear, baseline corrected response that can be used for quantitative work and it sharpens, or partially deconvolves, fused peaks in the chromatographic direction.

[00191] As described above, one output of the ion detection and quantitation method of embodiments of the present invention is a table or list of parameters corresponding to detected ions. This list can be described as a table with least three and possibly more columns. The three columns in each row in the table are used to store a detected ion's retention time, mass-to-charge ratio, and intensity. Additional columns can contain, for example, detected ion's width as measured by the FWHM or the zero-crossing of the convolved peak corresponding to the ion. The smoothing filter measures the FWHM of the peak and the second derivative filter measures the zero-crossing width. The number of rows in the table corresponds to the number of ions detected.

[00192] The present invention also provides a data compression benefit. This is because the computer memory needed to store the information contained in the ion table is significantly less than the memory needed to store original LC/MS data. For example, a typical injection that contains 3600 spectra (for example, spectra collected once per second for an hour), with 400,000 resolution elements in each spectrum (for example, 20,000:1 MS resolution, from 50 to 2,000 amu) requires in excess of several gigabytes of memory to store the LC/MS data matrix of intensities.

[00193] In a complex sample, using embodiments of the present invention on the order of 100,000 ions can be detected. In embodiments of the present invention, these detected ions are represented by a table having 100,000 rows, each row corresponding to a detected ion. Each row would have at least three entries corresponding to the ion's distinct retention time, m/z and intensity triplet. The amount of computer storage required to represent such a table is typically less than 100 megabytes. This

storage amount represents only several percent of the memory needed to store the original data. Thus, less storage is required to store the ion date in tabular form than compared to the that required to store the original LC/MS data matrix. Thus, the ion detection and quantitation technique pf embodiments of the present invention also produces significant reduction in data storage requirements to store the detected ion parameter data. The ion parameter data stored in the ion table can be accessed and extracted for further processing (described below). Other methods for storing the data can be employed in embodiments of the present invention.

[00194] Using the ion detection technique of embodiments of the present invention, the complexity of spectra resulting from an era LC/MS experiment can be reduced even further than by the reduction provided by LC separation. For example, an LC separation may reduce a sample having 1000 molecules to one having 10 molecules. Despite the low number of molecules in the flow cell, each molecule may be eluting at different retention times. For example, in the exemplary case of 10 molecules in the flow cell, three of the molecules may correspond to peaks that are just starting to elute from the column, four of the molecules may be in the flow-cell in the middle of their elution profile, and the final three molecules may correspond to the tail of peaks that are exiting the flow cell.

[00195] Biological samples are an important class of mixtures commonly analyzed using LC/MS method. Biological samples generally comprise complex molecules. A characteristic of such a complex molecule is that a singular molecular species may produce several ions. Peptides occur naturally at different isotopic states. Thus, peptides that appears at a given charge will appear at several values of m/z. With

sufficient resolution, the mass spectrum of a peptide exhibits a characteristic ion cluster.

[00196] Proteins, which typically have high mass, are ionized into different charge states. The isotopic variation in proteins may not be resolved by a mass spectrometer. However, ions that appear in different charge states can be resolved, again producing a characteristic pattern.

[00197] Mass spectrometers measure only the ratio of mass-to-charge, not mass by itself. It is possible however, to infer the charge state of molecules such as peptides and protein from the pattern of ions they produce. Using this inferred charge state, the mass of the molecule can be estimated. For example, if a protein occurs at multiple charge states, then it is possible from the spacing of m/z values to infer the charge, to calculate the mass of each ion knowing the charge, and ultimately to estimate the mass of the uncharged parent. Similarly, for peptides, where the m/z changes due to change in the isotopic value for a particular mass m, it is possible to infer the charge from the spacing between adjoining ions.

[00198] There are a number of well-known techniques that utilize the m/z values from ions to infer the charge and parent mass. A requirement for each of these techniques is selection of the correct ions and the use of accurate values for m/z. Ions contained in the detected ion table generated by embodiments of the present invention as described above provide high precision values that can be used as inputs to these techniques to produce results with enhanced precisions.. In addition, several of the cited methods attempt reduce the complexity of spectra by citing various techniques to distinguish between ions that may appear in a spectrum.

[00199] Generally, these techniques involve selecting a spectrum centered on a prominent peak or combing spectra associated with a single peak, to obtain a single extracted MS spectra. If that peak were from a molecule that produced multiple, time-coincident ions, the spectra would contain all those ions.

[00200] However, that spectrum will also typically contain ions from unrelated species. These unrelated species can be from ions that elute at the exact same retention time as the species of interest or, more commonly, the unrelated species are from ions that elute at different retention times. However, if these different retention times are within a window of approximately the FWHM of the chromatographic peak width, the ions from the front or tails of these peaks are likely to appear in the spectrum. The appearance of the peaks associated with unrelated species requires subsequent processing to detect and remove them. In some instances where they coincide, they may be biasing measurements.

[00201] Figure 16 illustrates an exemplary LC/MS data matrix that results from two parent molecules, and the resulting multiplicity of ions. In the example, the two species overlap chromatographically. Figure 16 also shows the resulting complex spectrum. These spectra were obtained by simply extracting the column spectra from the LC/MS data matrix. These spectra could also have been obtained from co-adding or combining spectra from the LC/MS data matrix.

[00202] To simply such couples spectra, embodiments of the present invention focus on a restricted retention time region. Although not all unrelated ions may be removed, by eliminating many, a more simplified and easier to interpret spectrum when compared to spectra produced by conventional methods is provided.

[00203] In both these cases, the ions appear in the LC/MS data matrix at exactly the same retention time. For peptides, those peptides that differ only by isotopic mass interact with the column identically and elute at a common retention time. Peptides and proteins that are ionized into different charge states elute at the same time. It is useful to be able to examine the ions from an LC/MS chromatogram and identify those ions that share a common retention time.

[00204] Embodiments of the present invention also provide a technique for obtaining simplified spectra from an LC/MS chromatogram. After the ion table is created as described above, ions can be selected in a narrow retention time window. The width of the window can be chosen to be no larger than the FWHM of the chromatographic peak. In some case smaller windows such as one tenth the FWHM of a peak are selected. For example, the ion having the highest intensity value can be selected. The retention time associated with this ion is noted. Then only those ions that are within a narrow window from this retention time are selected for inclusion in the spectrum. Those ions to be retained are determined by examining the retention times stored in the ion parameter table described above.

[00205] This window to includes all ions that elute nearly simultaneously, and are thus candidates for being related. Likewise, the window excludes molecules that cannot be related. Many of these molecules would have been included in spectra obtained using conventional techniques. Thus, the resulting spectra obtained from the peak list contains only the ions corresponding to the species of interest, and produces a significantly simplified spectrum by excluding ions that cannot be related to the species of interest, by virtue of their different retention time *i.e.,* having retention

times falling outside of the chosen window. Figure 16 shows exemplary simplified spectra that result from this targeted extraction used in the embodiment of the present invention.

[00206] As a sample is collected with an LC/MS system, it is preferred that a plurality of spectra be collected across the peak in order for the retention time to be accurately inferred. For example, in embodiments of the present invention 5 spectra per FWHM are collected.

[00207] It is possible to alternate the configuration of an LC/MS system on a spectrum by spectrum basis. For example, all even spectra can be collected as described above. All odd, interleaving spectra could be obtained with the MS in a different mode. One example of a mode change is provided by LC/MS/MS where fragmentation could be performed. Thus, the odd-numbered spectra that are collected are of the fragments of the ions collected in the un-fragmented state as shown by preceding even-numbered spectra.

[00208] In such a system, the fragment, or modified ions appear with a chromatographic profile having the same retention time as the unmodified ions. This is because the extra time required to perform modifications in online MS is short as compared to the peak width or FWHM of a chromatographic peak. For example, the transit time of a molecule in an MS is typically on the order of milli- or micro-seconds, while the width of a chromatographic peak is typically on the order of seconds or minutes.

[00209] The unmodified and modified spectra can be segregated into two independent data matrices. The operations of convolution, apex detection, parameter estimation

and thresholding described above can be applied independently to both. Although such analysis results in two lists of ions, the ions appearing in the lists bear a relationship to one another. For example, an intense ion having a high intensity that appears in the unmodified list of ions may have counterpart in the list of modified ions. In such a case, the ions will typically have a common retention time. To associate such related ions with one another for analysis, a window restricting retention time as described above can be applied to both data matrices. The result of applying such a window is to identify ions in the two lists having a common retention time and therefore likely to be related.

[00210] Figure 17 is a graphical chart illustrating how related ions can be identified in the unmodified and modified ion lists generated by an embodiment of the present invention. Data matrix 1702 shows three precursor ions 1704, 1706 and 1708 that are detected in spectra resulting from an unmodified MS experiment. Data matrix 1710 shows eight ions that result from an experiment after the MS is modified for example as described above to cause fragmentation. Ions in data matrix 1702 that are related to those in the data matrix 1710 appear at the same retention time, as indicated by the three vertical lines labeled t0, t1, and t2. For example, ions 1708a and 1708b in data matrix 1710 are related to ion 1708 in data matrix 1702. Ions 1706a, 1706b, and 1706c in data matrix 1710 are related to ion 1706 in data matrix 1702. Ions 1704a, 1704b and 1704c in data matrix 1710 are related to ion 1704 in data matrix 1702. These relationships can be identified by retention time windows with appropriate widths centered at t0, t1, and t2 respectively.

[00211] This process can be applied to more than two chromatograms, provided the retention time of peaks can be estimated from the data.

[00212] As described above, the present invention provides only values associated with each apex that are of interest. As a result, the convolved matrix can be further simplified into a simple, tabular list. Each row in the table corresponds to the properties of a single ion. For example, the table can be configured of such that the m/z value is in the first column, the retention time is in the second column and the intensity is in the third column. Other columns can be added to store other parameters associated with a peak that can be obtained from the convolved matrix, including, for example, characteristics of the shape of the convolved peak.

[00213] The resulting ion list or table can be interrogated to form novel and useful spectra. For example, as described above selection of ions from the table based upon the enhanced estimates of retention times produces spectra of greatly reduced complexity. These spectra have reduced complexity because use of the retention time window excludes ions unrelated to the species of interest as described above. Retention-time selected spectra simplify the interpretation of mass spectra of molecular species that induce multiple ions in a spectrum. Examples of molecular species that produce multiple ions at a common retention time are proteins, peptides, and their fragmentation products.

[00214] Similarly chromatograms of reduced complexity can be generated by basing the similarity of selected ions on common m/z values, that are found in the ion list.

[00215] Embodiments of the present invention also allow for analysis of peak purity that is whether the peak is due to a single ion or the result of co-eluting ions contained

in the peak. For example, by consulting the ion list generated by embodiments of the present invention and analyst can determine how many compounds or ions elute within the time of a principle peak of interest according to a figure of merit that describes the purity of a peak as follows:

[00216] The number of ions from the list occurring within a retention time window is summed. A ratio of the major ion is the peak ion having the highest intensity in the retention time window, to the sum is calculated. If the ration is 1:1 (100%), the peak is pure. If the ratio is less than 100%, the peak is not pure and there are coeluting ions in the peak. In addition, thresholding techniques can be used to determine a level of significance.

[00217] The foregoing disclosure of the preferred embodiments of the present invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many variations and modifications of the embodiments described herein will be apparent to one of ordinary skill in the art in light of the above disclosure. The scope of the invention is to be defined only by the claims appended hereto, and by their equivalents.

[00218] Further, in describing representative embodiments of the present invention, the specification may have presented the method and/or process of the present invention as a particular sequence of steps. However, to the extent that the method or process does not rely on the particular order of steps set forth herein, the method or process should not be limited to the particular sequence of steps described. As one of ordinary skill in the art would appreciate, other sequences of steps may be possible.

Therefore, the particular order of the steps set forth in the specification should not be construed as limitations on the claims. In addition, the claims directed to the method and/or process of the present invention should not be limited to the performance of their steps in the order written, and one skilled in the art can readily appreciate that the sequences may be varied and still remain within the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

1. A system for analyzing a sample, comprising:

a liquid chromatograph into which the sample is input for chromatographic separation;

a mass spectrometer to accept the output of the liquid chromatograph and outputs a plurality of spectra of the sample at discrete times;

a computer coupled to the mass spectrometer wherein the computer accepts the output spectra and stores them in a two-dimensional data matrix;

a two-dimensional filter;

wherein the computer applies the two-dimensional filter to the data matrix to generate an output data matrix and examines the output data matrix to detect ions in the sample by identifying one or more peaks in the output data matrix, wherein each peak corresponds to an ion in the sample.

2. The system recited in claim 1, wherein the data matrix is configured such that each column of the data matrix corresponds to a distinct one of the plurality of spectra at a discrete time and each row of the data matrix corresponds to a chromatogram of the sample for a particular mass-to-charge ratio.

3. The system recited in claim 1, wherein the peaks are detected by comparing each peak to a threshold and those peaks that exceed the threshold are deemed to be peaks associated with detected ions.

4. The system recited in claim 3, wherein the threshold is determined using a histogram of peak intensities.

5. The system recited in claim 1, wherein the filter is a matched filter.

6. The system recited in claim 1, wherein the filter is a rank-1 filter comprising a first filter that is convolved with the columns of the data matrix to generate a first intermediate matrix and

a second filter that is convolved with the rows of the intermediate matrix to generate the output data matrix.

7. The system recited in claim 6, wherein the rank-1 filter comprises one or more smoothing filters.

8. The system recited in claim 6, wherein the rank-1 filter comprises one or more second derivative filters.

9. The system recited in claim 1, wherein the filter is a rank-2 filter, comprising a first rank-1 filter and a second rank-1 filter, wherein the first rank-1 filter comprises a first filter that is convolved with the columns of the data matrix to generate a first intermediate matrix and a second filter that is convolved with the rows of the first intermediate matrix to generated a second intermediate matrix, and the second rank-1 filter comprises a first filter that is convolved with the columns of the data matrix to generate a third intermediate matrix and a second filter that is convolved with the rows of the third intermediate matrix to generate a fourth intermediate matrix and wherein the second and fourth intermediate matrices are combined to generate the output data matrix.

10. The system recited in claim 9, wherein the rank-2 filter comprises one or more smoothing filters.

11. The system recited in claim 9, wherein the rank-2 filter comprises one or more second derivative filters.

12. The system recited in claim 1, further comprising an ion list in which parameters corresponding to the ions are stored, wherein the parameters are determined by examining characteristics of the peaks in the output data matrix in an ion list.

13. The system recited in claim 12, wherein the each row of the ion list comprises one or more parameters associated with a particular ion in the sample to which the row corresponds.

14. The system recited in claim 13, wherein the one or more parameters comprise a mass-to-charge ratio associated with the particular ion, a retention time associated with the particular ion and an intensity associate with the particular ion.

15. The system recited in claim 14 wherein the one or more parameters comprise characteristics of the peak.

16. The system recited in claim 12, wherein the computer further produces a simplified spectrum or chromatogram by extracting related ions from the ion list to place in the simplified spectrum or chromatogram.

17. The system recited in claim 16, wherein the related ions are chosen as those ions falling within a retention window.

18. The system recited in claim 12, wherein the computer further produces a simplified chromatogram by extracting related ions from the ion list to place in the simplified spectrum.

19. The system recited in claim 18, wherein the related ions are chosen as those ions falling within a mass-to-charge window.

20. The system recited in claim 12, wherein one or more of the spectra are produced by modifying the mass spectrometer such that a set of spectra corresponding to the operation of the modified mass spectrometer are produced for analysis and a set of spectra corresponding to operation of the unmodified mass spectrometer are produced for analysis and a first ion list is generated for ions detected during operation of the unmodified mass spectrometer and a second ion list is generated for ions detected by operation of the modified mass spectrometer.

21. The system recited in claim 20, wherein related ions in the first and second ion lists are identified by applying a retention time window to the first and second ions lists.

22. The system recited in claim 20, wherein the modification is fragmentation switching.

23. The system recited in claim 1, wherein the two-dimensional filter is applied to the data matrix by convolving the data matrix with the two-dimensional filter.

24. A method for analyzing a sample, comprising:

introducing the sample into a liquid chromatograph for chromatographic separation to a liquid chromatograph output;

introducing the liquid chromatograph output into a mass spectrometer that outputs a plurality of mass spectra of the sample at discrete times;

inputting two or more of the plurality of mass spectra into a computer;

storing the two or more mass spectra in a two-dimensional data matrix;

specifying a two-dimensional filter to apply to the data matrix;

applying the two-dimensional filter to the data matrix to generate an output data matrix; and

examining the output data matrix to detect ions in the sample by identifying one or more peaks in the output data matrix, wherein each peak corresponds to an ion in the sample.

25. The method recited in claim 24, further comprising configuring the data matrix such that each column of the data matrix corresponds to a distinct one of the plurality of spectra at a discrete time and each row of the data matrix corresponds to a chromatogram of the sample for a particular mass-to-charge ratio.

26. The method recited in claim 24, further comprising:

comparing each peak to a detection threshold; and

identifying those peaks that those peaks that satisfy the detection threshold as peaks associated with detected ions.

27. The method recited in claim 26, further comprising:

creating a histogram of peak intensities from the data matrix; and

determining the detection threshold in accordance with the histogram.

28. The method recited in claim 24, wherein the two-dimensional filter is a matched filter.

29. The method recited in claim 24, further comprising:

specifying a rank-1 filter comprising a first filter and a second filter;

convolving the columns of the data matrix with the first filter to generate a first intermediate matrix; and

convolving the rows of the intermediate matrix with the second filter to generate the output data matrix.

30. The method recited in claim 29, wherein the rank-1 filter comprises one or more smoothing filters.

31. The method recited in claim 29, wherein the rank-1 filter comprises one or more second derivative filters.

32. The method recited in claim 24, further comprising:

specifying a rank-2 filter, comprising a first rank-1 filter and a second rank-1 filter, wherein the first rank-1 filter comprises a first filter and a second filter and the second rank-1 filter comprises a third filter and a fourth filter;

convolving the columns of the data matrix with the first filter to generate a first intermediate matrix;

convolving the rows of the first intermediate matrix with the second filter to generate a second intermediate matrix;

convolving the columns of the data matrix with the third filter to generate a third intermediate matrix;

convolving the rows of the third intermediate matrix with the fourth filter to generate a fourth intermediate matrix;

combining the second and fourth matrices to generate the output data matrix.

33. The method recited in claim 32, wherein the rank-2 filter comprises one or more smoothing filters.

34. The method recited in claim 32, wherein the rank-2 filter comprises one or more second derivative filters.

35. The method recited in claim 24, further comprising:

examining characteristics of the peaks identified as corresponding to detected ions to obtain parameters corresponding to the detected ions; and

storing the parameters corresponding to the detected ions in an ion list.

36. The method recited in claim 35, wherein the each row of the ion list comprises one or more parameters associated with a particular ion in the sample to which the row corresponds.

37. The method recited in claim 35, wherein the one or more parameters comprise a mass-to-charge ratio associated with the particular ion, a retention time associated with the particular ion and an intensity associate with the particular ion.

38. The method recited in claim 37 wherein the one or more parameters comprise characteristics of the peak.

39. The method recited in claim 35, further comprising extracting related ions from the ion list to create a simplified spectrum or chromatogram.

40. The method recited in claim 39, further comprising:

specifying a retention time window; and

identifying related ions from the ion parameter list as those ions having retention times falling within the retention time window.

41. The method recited in claim 39, wherein the computer further produces a simplified chromatogram by extracting related ions from the ion list to place in the simplified spectrum.

42. The method recited in claim 41, further comprising:

specifying a mass-to-charge ratio window; and

identifying related ions from the ion parameter list as those ions having mass-to-charge ratios falling within the mass-to-charge ratio window.

43. The method recited in claim 35, further comprising:

generating a set of spectra corresponding to the operation of the mass spectrometer for analysis;

storing a first ion parameter list for ions detected during operation of the mass spectrometer;

modifying the mass spectrometer;

generating a set of spectra corresponding to the operation of the modified mass spectrometer for analysis; and

storing a second ion parameter list for ions detected during operation of the modified mass spectrometer.

44. The method recited in claim 43, further comprising:

specifying a retention time window; and

identifying related ions from the first and second ion parameter list as those ions having retention times falling within the retention time window.

45. The method recited in claim 43, wherein the modification is fragmentation switching.

46. The method recited in claim 24, further comprising convolving the data matrix with the filter.

ABSTRACT OF THE DISCLOSURE

Chromatograms and mass spectra produced by an LC/MS system are analyzed by creating a two-dimensional data matrix of the spectral and chromatographic data. The two-dimensional matrix can be created by placing the spectra generated by the mass spectrometer portion of the LC/MS system in successive columns of the data matrix. In this way, the rows of the data matrix correspond to chromatographic data and the columns of the data matrix correspond to the spectra. A two-dimensional filter is specified and applied to the data matrix to enhance the ability of the system to detect peaks associated with ions. The two-dimensional filter is specified according to desired criteria. Rank-1 and rank-2 filters can be specified to improve computational efficiency. One method of applying the two-dimensional filter is through convolution of the data matrix with the two-dimensional filter to produce an output data matrix. Peaks corresponding to detected ions are identified in the output data matrix. Parameters of the peaks are determined and stored for later processing including simplification of chromatograms or spectra, by for example, identifying peaks associating with ions having retention times falling within a specified retention time window or having mass-to-charge ratios falling within a specified mass-to-charge ratio window.

APPARATUS AND METHOD FOR IDENTIFYING PEAKS IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY DATA AND FOR FORMING SPECTRA AND CHROMATOGRAMS

Marc Victor Gorenstein, Robert Stephen Plumb, Chris Lee Stumpf
Waters Corporation

This document contains additional material to be included with the provisional.

Filter embodiments

Apodized Savitzky-Golay filters

The row and column filters employed in the preferred embodiments are modifications to the well-known Savitzky-Golay (SG) filters. These modifications are original, and the resulting filters are termed: Apodized Savitzky-Golay (ASG) filters.

The following code (ANSI-C) returns the N filter coefficients (specified in the code as ncoef) of an Apodized Savitzky-Golay filters (ASG).

The calling function (defined in the code below) is

```
int ApodQuadFilterCoef (double *coef, int ncoef, int nderiv)
```

If the parameter nderiv = 0, the coefficients (returned in the array coef) are smoothing coefficients for an ASG filter. If the parameter nderiv = 2, the coefficients (returned in the array coef) are second derivative coefficients from an ASG filter.

The term apodization refers to filter coefficients that are obtained via applying an array of weight coefficients to the least-squares derivation of SG filter coefficients. The weight coefficients are the apodization function. For the ASG, the apodization function is a cosine window (defined by `COSINEWINDOW`). This apodiziation function is applied, via weighted least-squares to a box-car filter to obtain the ASG smoothing filter, and to a 2nd derivative SG quadratic polynomial, to obtain the ASG 2nd derivative filter. The box car filter and 2nd derivative quadratic are, by themselves, special cases of Savitzky-Golay polynomial filters.

Apodiziation preserves the smoothing and differentiation properties of SG filters, while producing a much improved high-frequency cutoff characteristics. Specifically, apodiziation removes sharp transitions of the SG filter coefficients at the filter boundaries, and replaces them with smooth transitions to zero. (It is the cosine apodization function that forces the smooth transition to zero.).

The column and row filters used in the preferred embodiment are these smoothing and 2nd derivative ASG filters.

```
/*****************************************************************************
TITLE:      ApodQuadFilterCoef

PURPOSE:  Returns Apodized Savitzky Golay filter coefficients
    for a quadratic polynomial model. The coefficients can extract
    from data a smoothed, first or second derivatives curve.

OPERATION:  Coefficients are calculated from normal equations.
    Design matrix for ncoef = 7 is
      1  -3  9/2
      1  -2  4/2
      1  -1  1/2
      1   0  0
      1   1  1/2
      1   2  4/2
      1   3  9/2

    Apodization is performed by a cosine window where
    weight =  [ 1+ cos(pi * ii/(half + 1)) ] /2
    so for ii=0,       weight = 1
       for ii=+/-(half+1),  weight = 0;

INPUT:    coef  pointer to array to which filter coefficients
          are written. User must allocate memory.
      ncoef the number of coefficients, which must be an
          an odd number >= 3.
      nderiv  = 0 smooth, = 1 first derivative, or
          = 2 for second derivative.

RETURNS:  ncoef Success. ncoef = the number of coef in coef.
      -1    Failure, which occurs if ncoef <3,
          or if nderiv is not equal to 0,1, or 2.

HISTORY:  June  1998, M. Gorenstein
COPYRIGHT (C) 1998 Waters Corp.
*****************************************************************************/
#define COSINEWINDOW(kk,nhalf) ( 1.0+cos(PI * (double)kk/(nhalf+1.0)) )
int ApodQuadFilterCoef (double *coef, int ncoef, int nderiv)
{
  int ii, nhalf;
  double c00=0.0, c11=0.0, c22=0.0, c02=0.0, det;
  double d0, d1, d2, weight;

  nhalf = (ncoef-1)/2;
  ncoef = nhalf*2+1;  /* Just in case ncoef is even */

  if (ncoef<3)    return(-1);

  /* Computation is complicated by c02 cross term */
  if (nderiv==0 || nderiv ==2 )
  {
    /* Elements of correlation matix */
    for (ii=-nhalf; ii<=nhalf; ii++)
    {
      weight  = COSINEWINDOW(ii,nhalf);
      d0    = 1.0;
      d2    = ii*ii/2.0;
```

```
      c00   += SQR(weight)*d0*d0;
      c02   += SQR(weight)*d0*d2;
      c22   += SQR(weight)*d2*d2;
    }

    det = c00*c22 - SQR(c02);

    /* 2 by 2 matrix inversion performed in each expression */
    for (ii = -nhalf; ii<=nhalf; ii++)
    {
      weight  = COSINEWINDOW(ii,nhalf);
      if (nderiv==0)
        coef[nhalf+ii] = SQR(weight)*(c22 - SQR(ii)*c02/2.0)/det;
      else
        coef[nhalf+ii] = SQR(weight)*(c00*SQR(ii)/2.0 - c02)/det;
    }

    return(ncoef);
  }
  else if (nderiv==1)
  {
    for (ii=1; ii<=nhalf; ii++)
    {
      weight  = COSINEWINDOW(ii,nhalf);
      d1    = ii;
      c11   += SQR(weight)*d1*d1;
    }
    c11 *= 2.0;

    for (ii= -nhalf; ii<=nhalf; ii++)
    {
      weight = COSINEWINDOW(ii,nhalf);
      coef[nhalf+ii] = SQR(weight)* ii / c11;
    }

    return(ncoef);
  }
  /* Illegal derivative number */
  return(-1);
}
```

R lationship between widths of smoothing and differentiating filters for rank-2 convolution filter.

A rank-2 filter is obtained by summing two rank-1 filters. Each rank-1 filter matrix is the outer matrix product of a column filter and a row filter.

The filter widths of each of the column filters (their coefficient number) are set in proportion to the spectral peak width.

The filter widths of each of the row filters (their coefficient number) are set in proportion to the chromatographic peak width.

In the preferred embodiment, the widths of the column filters are set equal to each other. That is, the width of each column filter is set equal to each other an in proportion to the spectral peak. Thus for example, for a spectral peak width FWHM of 5 channels, we may choose to set the filter width to 11 points, so the filters width of both the smoothing and 2^{nd} derivative spectral filter will be set to the same value of 11 points.

Analogously, in the preferred embodiment, the widths of the row filters are set equal to each other. That is, the width of each row filter is set equal to each other an in proportion to the chromatographic peak. Thus for example, for a chromatographic peak width FWHM of 5 channels, we may choose to set the filter width to 11 points, so the filters width of both the smoothing and 2^{nd} derivative spectral filter will be set to the same value of 11 points.

The result of this choice of widths is that the dimensions of the rank-1 filters are equal. That is, if the first rank-1 filter has dimension M X N, then the dimension of the second rank-2 filter has dimension M X N. This is not a requirement for the method, but it the filters of the preferred embodiment satisfy this rule.

Normalization of rank-1 filters used to construct a rank-2 convolution filter.

The rank-2 filter is obtained from the sum of two rank-1 filters. The resulting filter profile (also termed the finite-impulse response, or the point-source response) is then determined by the relative normalization of the two rank-1 filters.

$$F^1_{p,q} = f^1_p g^1_q$$
$$F^2_{p,q} = f^2_p g^2_q$$

The top equation is the point source response of the first rank-1 filter. The bottom equation is the point source response of the second rank-1 filter.

In the preferred embodiment, the row dimension of both rank-1 filters is the same, and the column dimensions of each rank-1 filter is the same, as noted in the above section. (The row dimension can be different from the column dimension). In this case, we can simply add the coefficients to obtain the rank-2 convolution filter's point source response. This sum is then

$$F_{p,q} = f^1_p g^1_q + f^2_p g^2_q$$

Clearly the relative normalization of the two rank-1 filters determines the overall points source response $F_{p,q}$. In the preferred embodiment, each rank-1 filter is normalized so that the sum of its coefficients squared equals one. That is

$$\sum_{q=1}^{N}\sum_{p=1}^{M}\left(f^1_p g^1_q\right)^2 = 1$$
$$\sum_{q=1}^{N}\sum_{p=1}^{M}\left(f^2_p g^2_q\right)^2 = 1$$

Here, we see that the matrix of coefficients of each rank-1 filter has dimension M by N. These two equations show that the sum-of-the-squares of each matrix element in each rank-1 filter sum to one.

The smoothing filters and 2nd derivative filters of the preferred embodiments can be normalized to satisfy this criteria by applying an appropriate scaling factor to the coefficients of the respective rank-1 matrices.

Extraction of local maxima of matched 2nd derivative filters for both chromatographic and spectra peak location.

The method that locates a single apex for each ion can also be used to locate apices in each row or column of the matrix. These apices may be useful to store a spectra or chromatograms at known times or mass values.

Spectra or chromatograms obtained from the second derivative filters can be obtained for each row and column. These intermediate results can be examined for local maxima as well. These maxima are, in effect smoothed versions of the chromatograms and spectra. Local maxima can be extracted and saved, giving additional detail as to the spectral content of the sample at a particular time or time range, or the chromatographic content at a typical mass or mass range.

Real-time embodiment of rank-1 and rank-2 filters.

The rank-1 and rank-2 filter formulation lend themselves to a real-time formulation of filtering a matrix of data.

In a convention LC/MS system, spectra are acquired as the separation progresses. Typically spectra are written to computer memory at a constant sample rate (typical value is once per second). From there, the spectra are written to more permanent storage, such as to a hard disk memory.

One embodiment of the method is to obtain the convolution matrix only after the acquisition is complete. Thus the original data can be preserved, and the convolved matrix itself can be preserved, as well as the ion list obtained from the local maximum.

Another embodiment is to obtain the columns of the convolution matrix while the data is being acquired. Thus the initial columns can be obtained, analyzed, and have their ions written to disk before the acquisition of spectra is complete.

This real-time embodiment of the method essentially analyzes the data in computer memory, writing only the ion list to the permanent hard disk drive. By real time, we mean that the rank-1 and rank-2 processing is performed on the spectra in computer memory as the data is being obtained. Thus the ions detected by the LC/MS in the beginning the separation are detected in the spectra written to disk by this convolution method and the portion of the ion list containing the ions is also written to disk as the separation proceeds.

There are time delays involved. The spectra containing ions elute in a chromatographic peak at time t, and width, delta t, can be processed as soon as they are collected. Typically the processing can then begin at t + 3 delta t. The ions from this peak are then written to the computer disk.

The implementation of the algorithm parallels what has been described in the body of the text. The results of the real time implementation are identical to that would be obtained by a post-processing embodiment.

The advantage to real-time processing is that

1) the ions list is obtained quickly
2) The information in the ion list can be used to trigger other real-time processes
3) Other real time processes that can be triggered by obtaining the ion list in real time include fraction collect, or stop flow technique to store eluent for analysis.
4) Example of stop-flow technique are those where the elutent is trapped in a nuclear-magnetic-resonance (NMR) spectral detector.

A particularly efficient real-time formulation, that is the preferred embodiment, is to replace each non-zero element of a scan at it arrives, by the filter coefficient scaled by the element intensity. The scaled filter coefficients are then added to a spectral buffer.

The spectral buffer is an array. The number of elements in the spectral buffer equals the number of elements in each spectrum. When each non-zero scan element arrives during a spectrum scan, the scaled filter coefficients are added to the spectral buffer. The center of the filter coefficients is located to correspond to the element in the spectral buffer corresponding to the scan element whose intensity was just received.

In a real-time formulation the original spectrum need never to be recorded to computer memory. Only the filtered scan is recorded. For the rank-1 formulation, only a single spectrum buffer is needed. For the rank-2 formulation, two spectral buffers are needed, one for the smoothing, and one for the 2^{nd} derivative spectral filters.

Additional storage memory is need for the real time formulation. For the rank-1 filter, at the end of each scan, the spectral buffer, containing the filter spectrum has to be added to a chromatographic buffer. The chromatographic buffer contains N-spectra, where N is the number of coefficients in the chromatographic buffer.

This chromatographic buffer is a FILO, first in last out buffer. When a new spectrum is added, the oldest spectrum is dropped. When a new spectrum is added, the chromatographic filter is applied to each row of the chromatographic buffer. The output of this filter is a single column of the convolution matrix.

These single columns are themselves added to a apex buffer. The apex buffer is three spectra width and each column is the length of a complete spectrum. This is also a FILO buffer. Each column is a column from the convolved matrix. When a new column is added, the oldest is dropped. The local maxima of the central column are recorded as the ions. That is, the local maximum intensity and the interpolation in retention time and mass to provide accurate retention time and mass values are obtained from this three spectrum apex buffer. Spectral peak width information is obtained by examining points adjacent to the local maxima along the column.

This three-spectrum FILO apex buffer can of course be expanded. To measure chromatographic peak width from the convolved data, it would be necessary to expand the apex buffer to include the number of spectra at least equal to the FWHM of the chromatographic peak. In a preferred embodiment, the number of convolved spectra in the apex buffer would correspond to twice the FWHM of the chromatographic peak.

Changing the filter characteristic via a schedul .

Filter characteristics such as the filter width and the scaling of the filters can be changed in response to the known changing characteristics of the LC separation or of the MS scans.

In a time of flight (TOF) MS, the peak width is known to change from low values (such as 0.010 amu) to wider values (such as 0.130 amu) over the coarse of each scan. This changing resolution as a function of m/z is a well-known and fundamental property of TOF mass spectrometers.

The width of the spectral filters, both smoothing and differentiating, is described by their coefficient number. This coefficient number is, in the preferred embodiment, set equal to about twice the width of the mass spectrometric peak, where width is the full width of the peak measured at half height (Full width at half maximum or FWHM). As the MS scan progresses, say from low to high mass, the filter width of both the smoothing and 2nd derivative column filters employed by the preferred embodiment can be expanded accordingly to preserve the relationship between filter width and peak width.

Analogously, if the width of the chromatographic peak is known to change during a separation, the width of the row filters can themselves be expanded or contracted to preserve the relationship between filter width and peak width

Measurements of peak width

From the convolved matrix, the width a peak in the spectral direction can be obtained by locating the nearest zero crossing points that straddle the apex. The distance between the zero-crossing is a measure of spectral peak width.

From the convolved matrix, the width a peak in the spectral direction can be obtained by locating the nearest minima that straddle the apex. The distance between the minima is a measure of spectral peak width.

From the convolved matrix, the width a peak in the chromatographic direction can be obtained by locating the nearest zero crossing points that straddle the apex. The distance between the zero-crossing is a measure of chromatographic peak width.

From the convolved matrix, the width a peak in the chromatographic direction can be obtained by locating the nearest minima that straddle the apex. The distance between the minima is a measure of chromatographic peak width.

The advantage of measuring spectral or chromatographic peak width is that these numbers can be used to confirm that a peak is resolved from its neighbors. If a large value of peak width can be used to flag peaks that may be coincident. The locations of zero crossings or local minimum can be used as input to estimate the effect of interfering coincidence or to in fact modify the values found in the ion list.

Changing or attenuated MS intensity

If the attenuation of the mass spectrometer in intentionally or inadvertently change, the spectral column filters can be scaled (by multiplying their filter coefficients by a scaling factor) to compensate for the change.

Further, if value for elements are known to be invalid (say that result from detector saturation) than these values can be removed or modified or edited prior to the filtering steps.

Mass interpolation embodiment

The coefficients of the filters can be modified to produce interpolated results to take into account possible small changes due to the mass calibration of the instrument. These changes can be made from spectrum to spectrum. That is if a change in mass calibration occurs that corresponds to an offset of a fraction of a channel, say 0.3, then the column filters (both smoothed and 2^{nd} derivative) can be derived that in effect estimate what the output would be in the absence of such a mass offset. Thus a real-time mass correction can be made. The resulting filter will by slightly asymmetric in order to account for this offset.

Uses of table of the method

Finger printing or mapping

There are many examples of mixtures that are, on the whole well characterized, and have essentially the same composition, and whose components exist in the same relative amounts. Biological examples include the end products of metabolism such as urine, cerebrospinal fluid, and tears. Other examples are the protein contents of cell populations found in tissues and blood. Examples in industry include perfumes, fragrances, flavors, fuel analysis of gasoline or oils.

Examples of variations from the norm in these fluids are xenobiotics in the case of products of metabolism that result from ingestion or injection of drugs or drug substances; evidence of drugs of abuse in metabolic fluids; adulteration in products such as juices, flavors, and fragrances; or in fuel analysis.

The list of ions that can be obtained from the method disclosed here can serve as a input to methods known in the art for fingerprint analysis. Packages such as SIMCA (Umetrics, Sweeden), or Pirouette (Infometrix, Woodenville, Washington, USA) can take as input the list of ions produced by this method and reveal changes in ions between sample populations.

These analyses can determine the normal distribution of entities in a mixture, and then identify those samples that deviation from the norm.

Synthetic route

The synthesis of a compound may produce the desired compound together with additional molecular entities. These additional entities characterize the synthetic route. The ion list can be a finger print that can be used to characterize the synthetic route of the synthesis of a compound.

Biomarker discovery

Another important application of this method is to biomarker discovery. The discovery of molecules whose change in concentration correlates uniquely with a disease condition or with the action of a drug is fundamental to the detection of disease or to the processes of drug discovery.

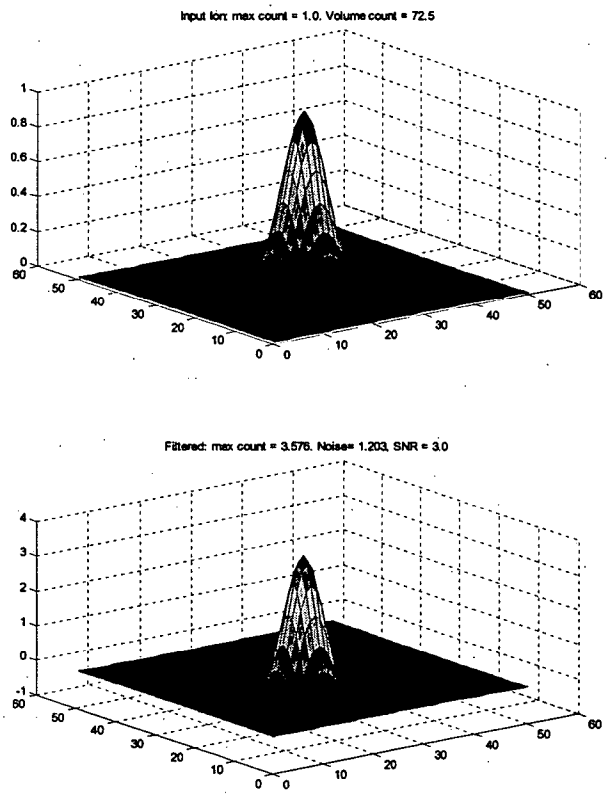
Biomarker molecules can occur in cell populations or in the products of metabolism or in fluids such as blood and serum. Comparison of the ion list obtained control and disease or dosed states by methods known in the art (cited above) can be used to identify molecules that are markers for the disease or for the action of a drug.

Reduction in computing tim

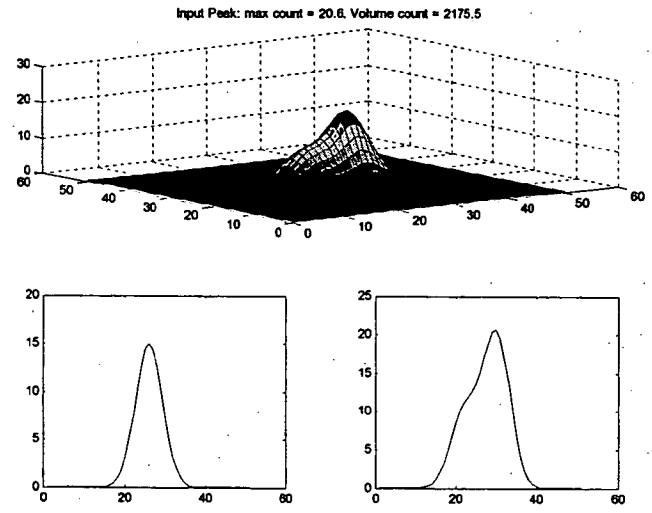
All analyses of data obtained from LC/MS system is speeded up if the analysis is performed on the ion list than on the original data. The original continuum data obtained from an LC/MS experiment can contains 200x200,000 elements or 40,000,000 sampled points. The ion list from a complex sample contains 200,000 ions.

Shoulder detection

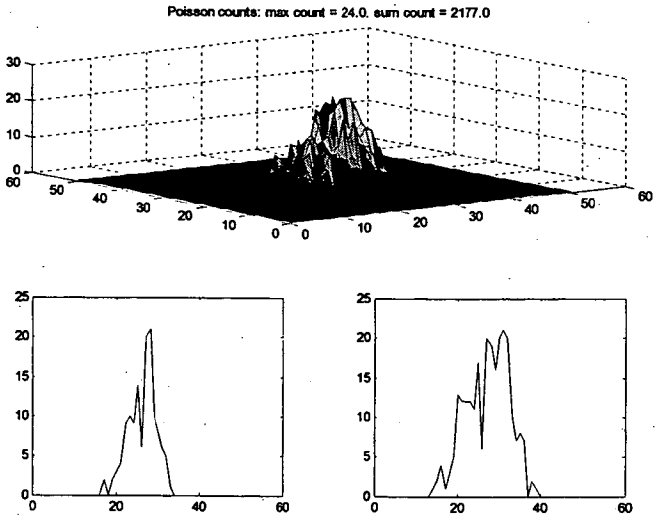
The rank-2 convolution filter of the preferred method contains a 2nd derivative filter in both the chromatographic and spectral directions. This filter can detect shouldered peaks. A shouldered peak occurs when a peak of low intensity is nearly coincident with a peak of higher intensity. The apex of the lower peak may not be evident in the data. Given that the rank-2 filter contains a 2nd derivative filter which measures curvature, the apex of the second peak, which is not seen in the data directly, can be detected as a separate apex in the convolution output matrix.



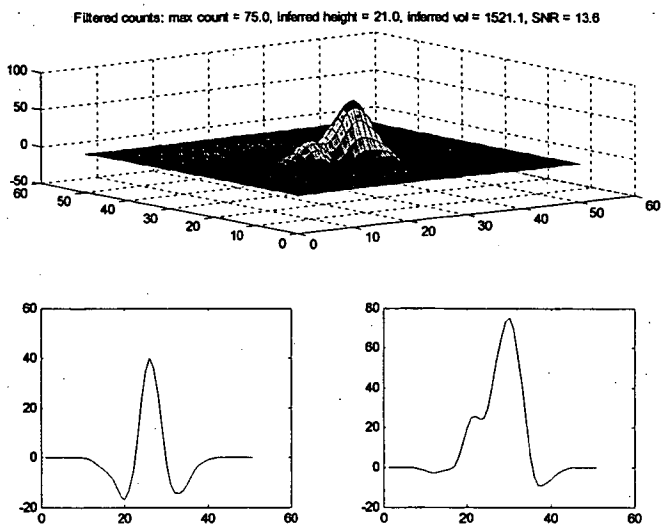
The top figure is a simulation of an LC/MS peak. The horizontal axis is time scan, the vertical axis is mass channel. The bottom figure is the point source response or finite impulse response of the rank-2 filter of the preferred embodiment.



The above figure is a simulation of two LC/MS peaks that have the same mass, and are nearly coincident in time. The result is a pure peak cross section in mass and a shoulder in time.



The top figure shows the effect of (simulated) counting noise (also termed shot noise) on the amplitude of each sampled element. The statistics of counting noise are described by Poisson statistics. The lower plots show the modified cross-section. Note the many local maxima that are produced as a result of the counting noise. Even though there are only two ions, the effect of counting noise is to produce many spurious local maxima.



The above figures show the convolution matrix that results from convolving the rank-2 filter with the simulated data. The resultant convolution matrix contains two distinct apices as can be seen in the top plot.

The two bottom plots show the cross-section. The lower right plot shows that two local maxima are seen in the chromatographic direction.

Thus the rank-2 convolution filter of the proposed method reduces the effect of counting noise, and deconvolves the shoulders to produce two local maxima. Each local maximum is associated with an ion. The properties of the ion, mass/charge, retention time and intensity can be obtained from the properties of the local maxima as described in the method.

Method for peak purity

The method obtains the peak purity by summing the intensities of all ions that elute within a retention time range of interest. Generally the retention time interval will correspond to the times between lift-off and touch-down of a peak of interest. The peak purity is then defined as

purity = 100 x (intensity of peak of interest)/(sum of intensity of all peaks in time range).

A more empirical formulation is

purity = 100 x (intensity of most intense peak)/(of intensity of all peaks in time range).

In these definition purity is a percent.

Ion Detection in Three Dimensions: A Novel Algorithm to Detect and Quantify Ions Obtained from High-Accuracy LC/MS Separations of Tryptic Digests of Complex Protein Mixtures

Introduction: The potential of LC/MS separations of complex mixtures is fully realized only when all the ions detected by the mass spectrometer are recovered in the analysis of the data. Once detected, the ions can be used for quantitative and qualitative purposes. Ions from isotopes of peptides, for example, can assembled into clusters and their mono-isotope mass can be accurately determined.

Thus the deceptively simple problem of ion detection is in fact, a potential limiting step in the exploitation of LC/MS data. For example, a peak-detection algorithm originally designed to detect peaks in a spectrum may be adopted to address the problem of ion detection in three-dimensional LC/MS separations, resulting in less than optimal performance. Here, we introduce a novel three-dimensional ion detection algorithm optimized for the analysis of high-mass-accuracy LC/MS data.

Methods: The method assembles the spectra obtained from the LC/MS separation into a matrix. The columns of the matrix are the spectra, the rows are the chromatograms. A novel convolution method, based on the properties of matched filters is applied to this matrix. The properties of the filter are designed to identify all the potentially detectable ions present in the data. Thus the approach lends itself to resolution enhancement: Pairs of ions that are only partially resolved or appear as shoulders, can separately quantified. In addition, low-intensity ions that might otherwise be overlooked can be detected; thus the method lends itself to the analysis of samples whose intensities span the full dynamic range of the instrument.

Results: The samples used to evaluate this new algorithm were obtained from tryptic digests of proteins test mixtures and from serum spiked with the test mixtures. We obtained data from these digests using a high-resolution (>17,000) orthogonal quadrupole time of flight mass spectrometers. The ions resulting from different isotopic states are separately detected and high accuracy mass, retention time, and intensity values for each ion are obtained. These cluster-associated ions are assembled into clusters producing unique value for mwHPlus for each cluster.

We demonstrate the quantitative reproducibility of molecular weight, retention time, and intensity of the data over the large dynamic range of this data as obtained using this algorithm.

Statistical study of LC/MS/MS data of human serum

This work provides a statistical study of LC/MS/MS data of human serum. The statistical study is very important for understanding the experimental data of complex biological mixtures. The digested peptides from sample (human serum) are run by LC/MS/MS. The raw data from LC/MS/MS are processed to generated ion sticks. Each ion stick has three parameters: m/z, retention time and intensity. One dimensional histograms of m/z, retention time and intensity for both MS data and MS/MS data are studied. Two dimensional histogram of m/z, retention time for both MS and MS/MS data are also studied. By those studies we can find what are most frequent m/z, retention time and intensity. Then the ion sticks are deconvoluted into peptide lists by both charge and isotopic deconvolution for both MS data and MS/MS data. Each peptide is come from multiple isotopes and multiple charges. Each peptide has three parameters: peptide mhp, peptide retention time and peptide intensity. The histograms of peptide mhp, peptide retention time and peptide intensity are studied. The most frequent peptide mhp, peptide retention time and peptide intensity are found. Two dimensional histogram of peptide's mhp, retention time for both MS and MS/MS data are also studied for different number of bins of mhp and retention time. This indicates how many peptides can be found in a certain mhp and retention time window for the complex biological mixtures.

Next step is to study the replication of injections. Sample (Human serum) is run by LC/MS/MS for replicated three times. The statistical calculations (mean, median, stand deviation and coefficient of variation) for number of peptides and total peptide intensities from 3 injections of the sample are provided. The coefficient of variation of number of peptides and total peptide intensities between 3 injections of the sample is about 5%. This demonstrates the reproducibility of total number of peptides and total peptide intensities. In summary we have studied the statistics of LC/MS/MS data of human serum, which is very useful for understating the experimental data of LC/MS/MS of complex biological mixtures.

Statistical study of LC/MS data of human serum spiked with five proteins

This work provides a statistical study of LC/MS data of human serum spiked with five proteins. The statistical study is an important step for quantitatively compare the relative level of proteins contained in two or more complex biological mixtures. Two samples are used for this study: sample 1 has human serum spiked with 5 pmole five proteins, sample 2 has human serum spiked with 1 pmole five proteins. The digested peptides from samples are run by LC/MS. Each sample has 3 replicated LC/MS runs. The raw data from LC/MS are processed to generated ion sticks. Each ion stick has three parameters: m/z, retention time and intensity. Then the ion sticks are deconvoluted into peptide sticks by both charge and isotopic deconvolution. Each peptide is come from multiple isotopes and multiple charges. Each peptide has three parameters: peptide mhp, peptide retention time and peptide intensity. The statistical calculations (mean, median, stand deviation and coefficient of variation) for number of peptides and total peptide intensities from 3 injections of each sample are provided. This demonstrates the reproducibility of total number of peptides and total peptide intensities.

Next step is to study the replication of each peptide from 3 injections of each sample. Number of replicated peptides and replicated intensities are studied. About 60% of peptides and about 90% of peptides intensities are replicated. This indicates the non-replicated peptides are small intensity one. The average of coefficient of variation of replicated intensities is about 20%. Then the replications of each peptide from 6 injections of two samples are studied. For all the peptides which are replicated for 6 times, the statistical calculations (mean, median, stand deviation and coefficient of variation) of peptide intensities from 3 injections of each sample are provided. The mean intensities of replicated peptides between two samples are compared to indicate the relative level change of spiked proteins contained in two samples. The ratio of mean intensities of replicated peptides between two samples is plot against mean coefficient of variation of intensities of two samples. In summary we have done statistical study of LC/MS data, which is very useful for quantitatively compare of the relative level of proteins contained in two or more complex biological mixtures.

Statistical study of LC/MS/MS data of human serum

This work provides a statistical study of LC/MS/MS data of human serum. The statistical study is very important for understanding the experimental data of complex biological mixtures. We study two cases: case 1 for one injection of one sample, case 2 for three or more injections of one or two samples.

Case 1 studies one injection of one sample: This study provides the statistics of ions and peptides of LC/MS/MS data of human serum. The digested peptides from human serum are run by LC/MS/MS for one time. The raw data from LC/MS/MS are processed to generated ion sticks. Each ion stick has three parameters: m/z, retention time and intensity. One dimensional histograms of m/z, retention time and intensity for both MS data and MS/MS data are studied. Two dimensional histogram of m/z, retention time for both MS and MS/MS data are also studied. By those studies we can find what are most frequent m/z, retention time and intensity. Then the ion sticks are deconvoluted into peptide lists by both charge and isotopic deconvolution for both MS data and MS/MS data. Each peptide is come from multiple isotopes and multiple charges. Each peptide has three parameters: peptide mhp (peptide mass plus proton mass), peptide retention time and peptide intensity. The histograms of peptide mhp, peptide retention time and peptide intensity are studied. The most frequent peptide mhp, peptide retention time and peptide intensity are found. Two dimensional histogram of peptide's mhp, retention time for both MS and MS/MS data are also studied for different number of bins of mhp and retention time. This indicates how many peptides can be found in a certain mhp and retention time window for the complex biological mixtures.

Case 2 studies three or more injections of one or two samples: This study provides the statistics of LC/MS replicated data of one or two samples. The statistical calculations (mean, median, stand deviation and coefficient of variation) for number of peptides and total peptide intensities from 3 injections of the sample (human serum) are provided. About 60% of peptides and about 90% of peptides intensities are replicated. For all the replicated peptides, histograms of mhp difference, retention time difference and intensity difference of replicated peptide's pair are studied. The statistical calculations (mean, stand deviation and coefficient of variation) for mhp and intensity of replicated peptides are also provided. Histograms of mean intensity of replicated peptides and non-replicated peptides are also studied. Similar statistical study for six injections of two samples (sample 1 has human serum spiked with 5 pmole five proteins, sample 2 has human serum spiked with 1 pmole five proteins) is also provided. The mean intensities of replicated peptides between two samples are compared to indicate the relative level change of spiked proteins contained in two samples.

Towards Quantitative Global Proteomics: Statistical Results Obtained from Multiple Tryptic Digests of Complex Protein Mixtures Using Novel Algorithms for the Detection, Tracking and Quantitation of Peptides

Introduction: Quantitation of proteins by high-mass accuracy LC/MS separations requires reproducible sample preparation, robust separation methods, and accurate mass measurements. With such high quality such data in hand, our attention must turn to the algorithms needed to extract information from this data. One critical algorithmic step is reliable tracking of molecular entities between samples.

A molecular entity detected in one injection could be located (i.e, tracked) in another injection by comparing only mass values. However, in the case of complex mixtures, such as tryptic digests, a retention-time search-window of a few minutes may contain pairs of entities that have the same *measured* mass, but in fact are unrelated. The resulting mistakes in tracking will compromise quantitation.

Methods: The novel algorithmic method introduced here addresses the problem of tracking. The method relies on accurate mass measurements to find the subset of entities that can be uniquely tracked by accurate mass alone. These unique matched pairs determine a retention time map, and such a map is found for all injections in a sample set.

These maps are then used to assign a unique reference retention time to *all* molecular entities in *all* injections. The method used the unique paired masses as, in effect, internal standards to correct for the retention time offset of all entities. The reference retention times of an entity can then be compared between any two samples in the sample set.

Results: The reference retention time puts all samples on an equal footing. The search window associated with the reference retention time can be as low as +/-0.2 minutes, much smaller than conventional minutes wide search windows. The reference retention time together with accurate mass can then be used to track an entity from injection to injection in a sample set.

Tryptic digests that contain upwards of 10,000 unique masses whose nearly 100,000 ions can be detected in a 2-hour LC separation followed by online MS detection.

4) TITLE:

Protocols to Assure Repr ducible Quantitative and Qualitative Analysis of Tryptic Digests of Complex Protein Mixtures for Global Proteomic Experiments

Introduction: Meaningful results in qualitative and quantitative proteomics, such as observation of differing expression levels of a protein in a series of samples, can only be obtained if samples are consistently prepared and analyzed. Tryptic digestion must be carried to completion for all proteins in order to maximize sequence coverage for identification and to all meaningful quantitative sample-to-sample comparison of a given peptide. Chromatographic separation of the resulting mixtures must also be performed in a consistent manner.

We have developed protocols for tryptic digestion of protein mixtures designed to assure reproducible peptide production, protocols to assure maximum reproducibility of capillary scale HPLC, and software tools to easily verify the reproducibility of our experiments.

Methods: A series of replicate digests of commercial rat serum was prepared. A proprietary detergent (RapiGest™ SF, Waters Corporation) was used as a denaturating agent. One or more standardized tryptic digests of individual proteins (MassPrep™ Digestion Standards, Waters Corporation) were added to the digests. Samples were analyzed by direction onto a 300 micron diameter x 15 cm column packed with Atlantis™ dC_{18} packing and eluted with a water/acetonitrile/formic acid gradient. The column effluent was directed a Nano Lockspray source on a hybrid quadrupole-time of flight mass spectrometer (Q-ToF Ultima API, Waters Corporation) Mass spectral data was obtained alternating scans of low and high collision cell energy. Every 10 seconds a separate reference sample spectrum was obtained.

Results: Use of the detergent as a denaturating agent was found not to interfere with chromatography or ionization of the tryptic peptides, nor was there any observable fouling of the ion source.

Sample consistency was demonstrated as follows: Raw mass spectral data was processed by Protein Lynx Global Server (Waters Corporation) to compile a list of data points as pairs of retention times and accurate mass values (observed m/z values at that moment corrected by use of the reference mass channel, accurate to 10 ppm or less). The resulting data are compared by submission to a software tool (Track 3D, Waters Corporation, patent pending) which correlates retention time, accurate mass values, and signal intensities of two or more samples. Results of this correlation show that signals for a given mass are observed at similar retention time from sample to sample for a great plurality of the observed signals as demonstrated on a graphical representation of difference in retention time vs. retention time for any pair of data sets. Furthermore, we observe that data that replicates in such a fashion represents a very high percentage of the total ion signal intensity for all the data in question, thus demonstrating reproducibility from sample to sample.

Fuller details of our protocols will be included in the poster.

Figures

Figure 1. The LC/MS system, showing the LC and the MS.

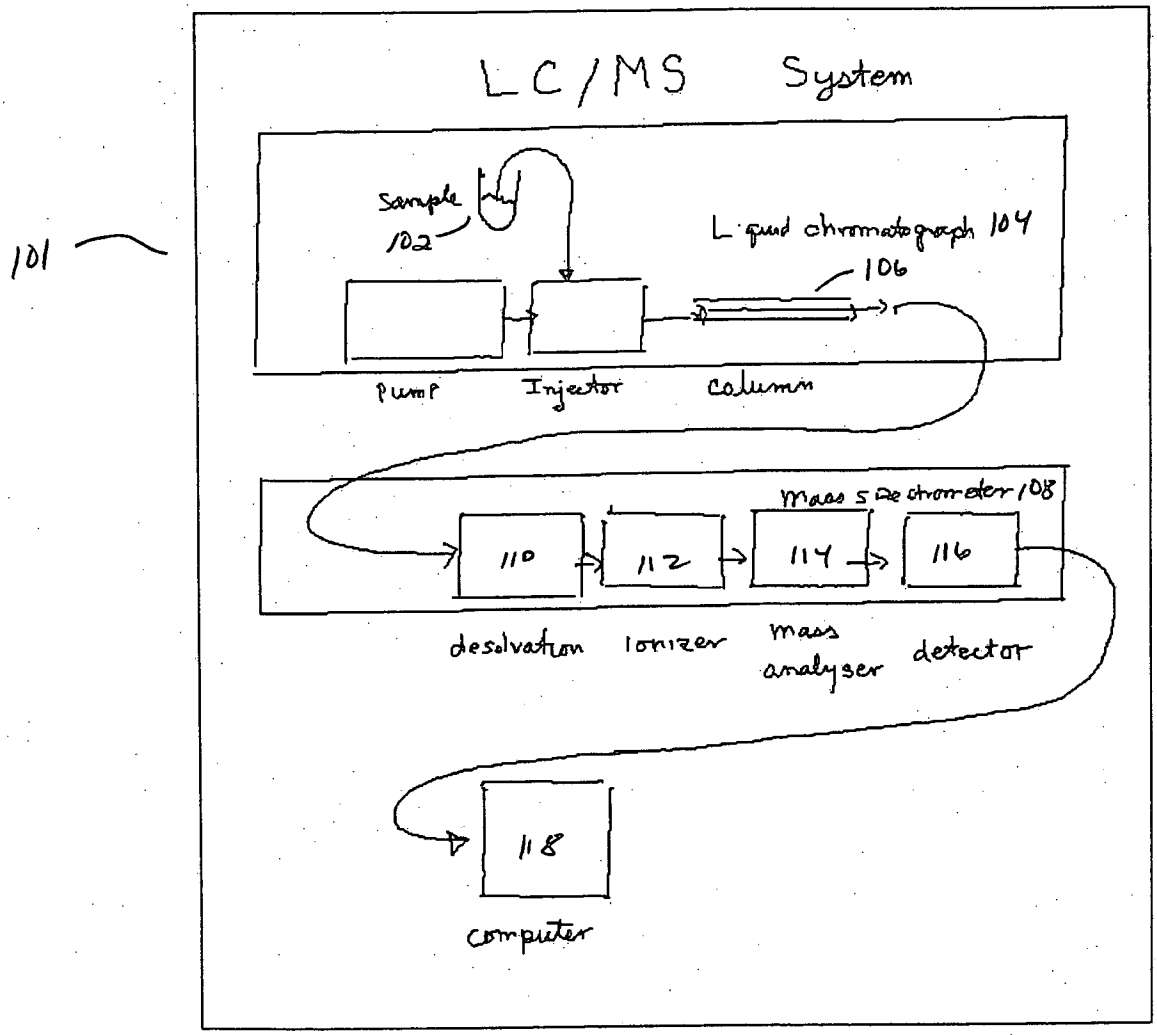


Figure 3 Three successive spectra, collected in tim .

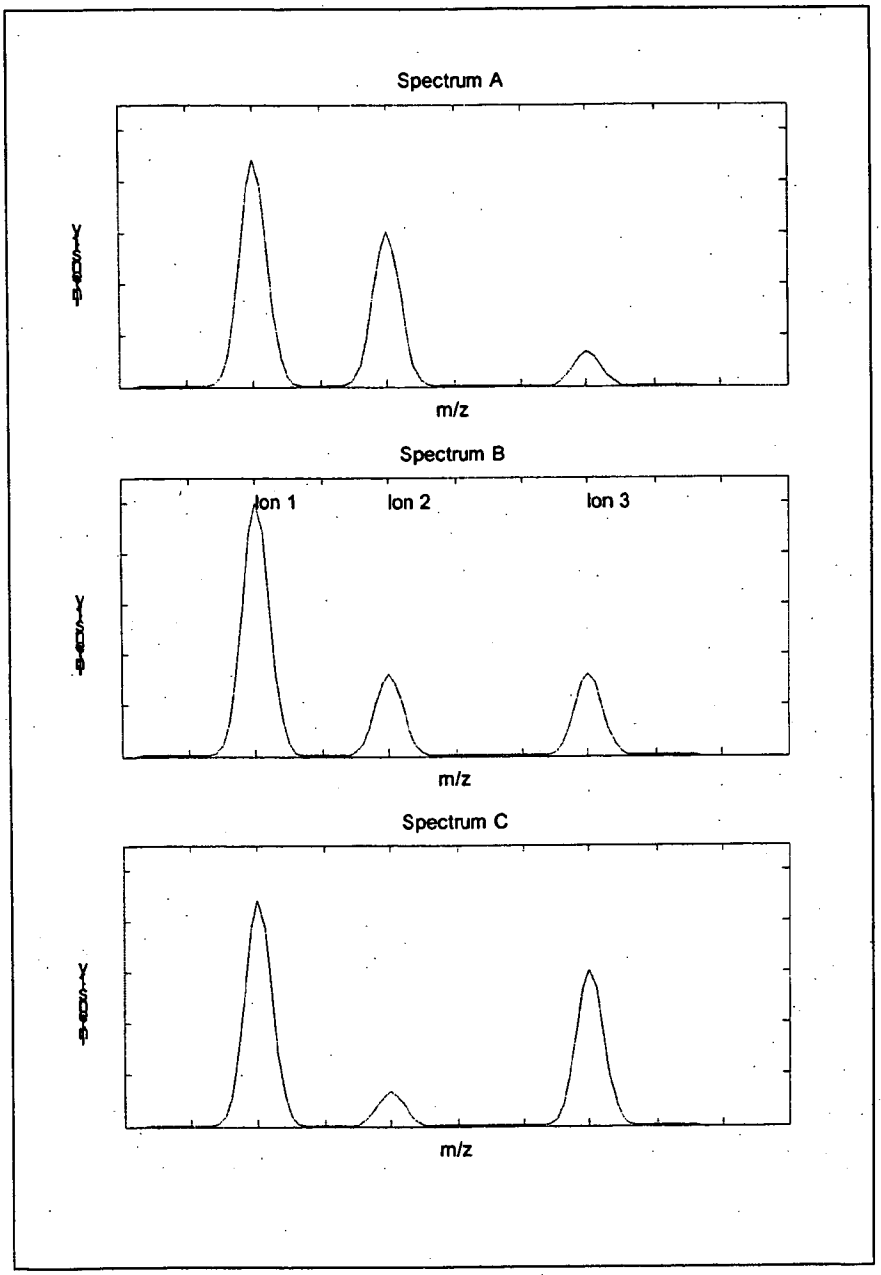
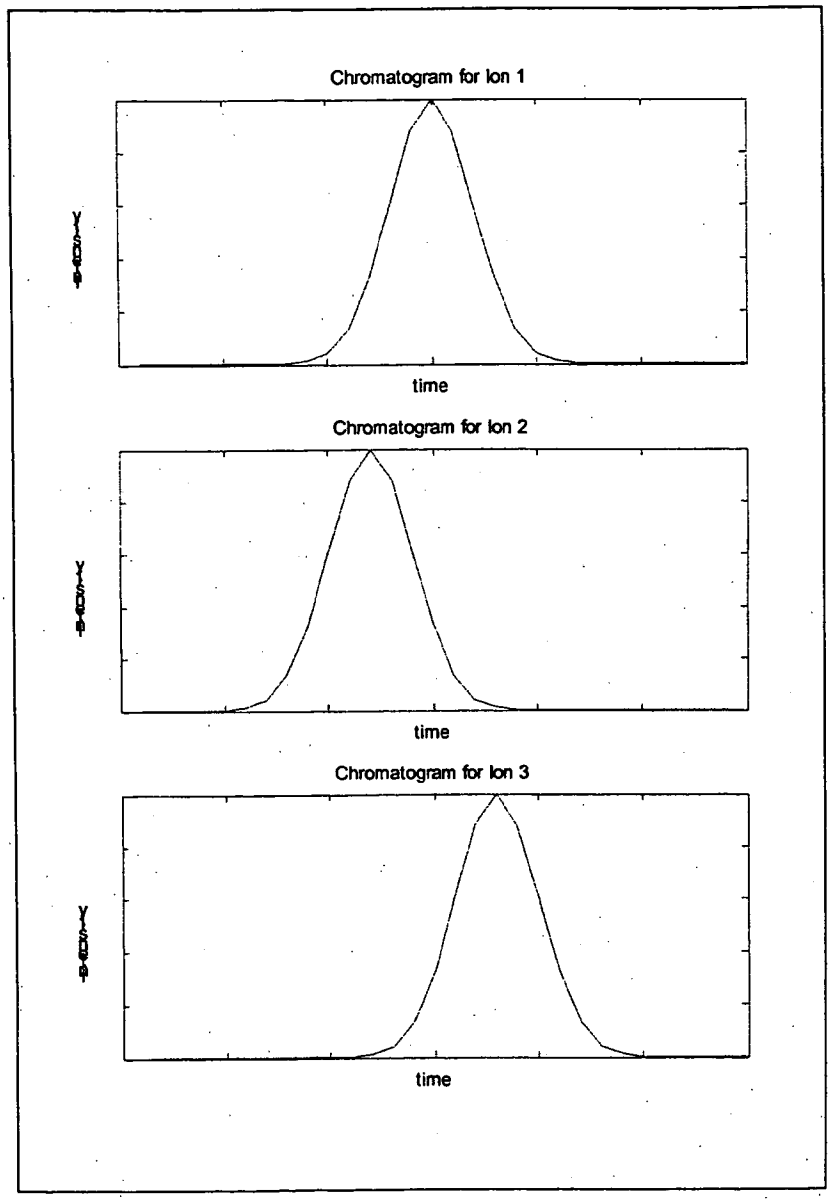


Figure 4 Chromatograms for three ions.



Figur_ 5 Contour plot of simulated LC/MS data matrix

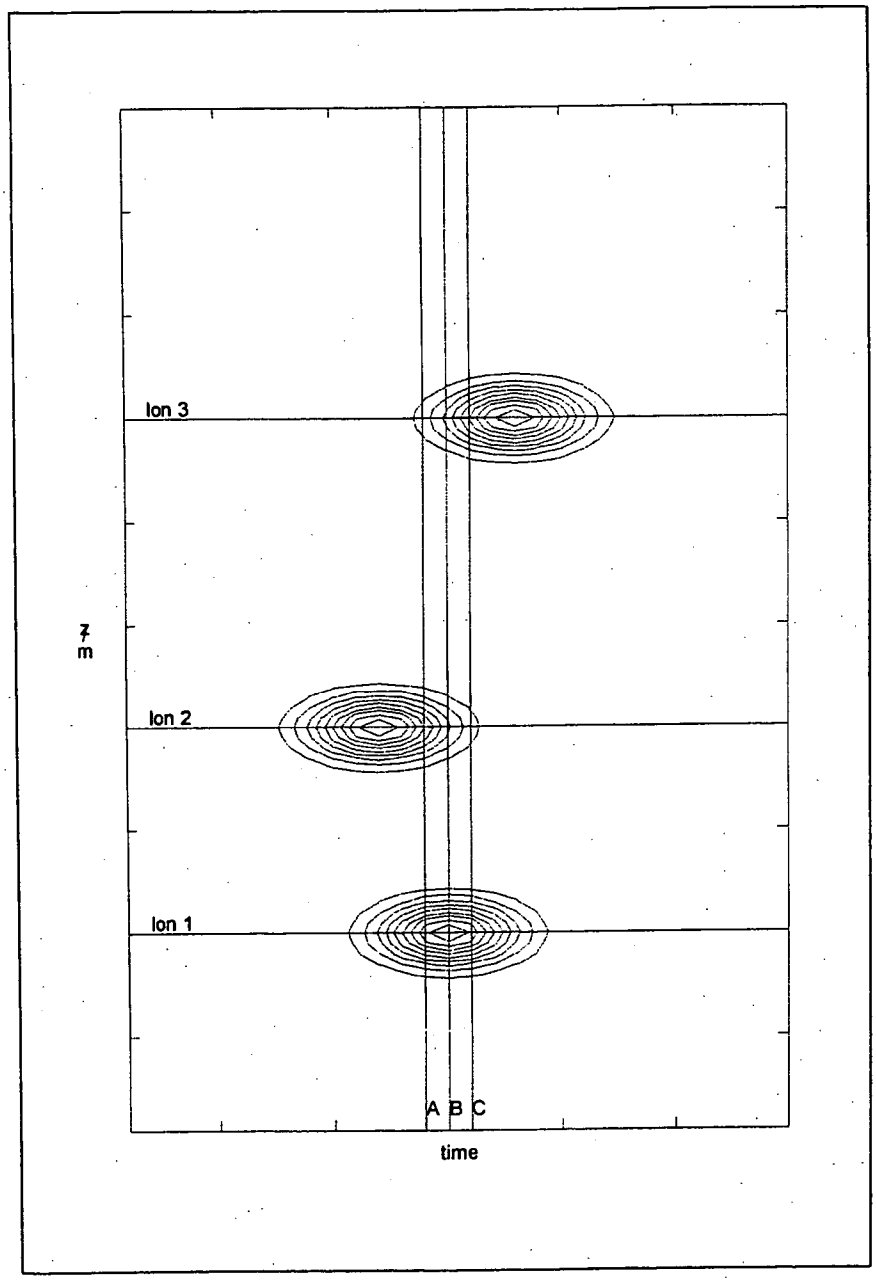


Figure 6. Example of coeluted ion in contour plot

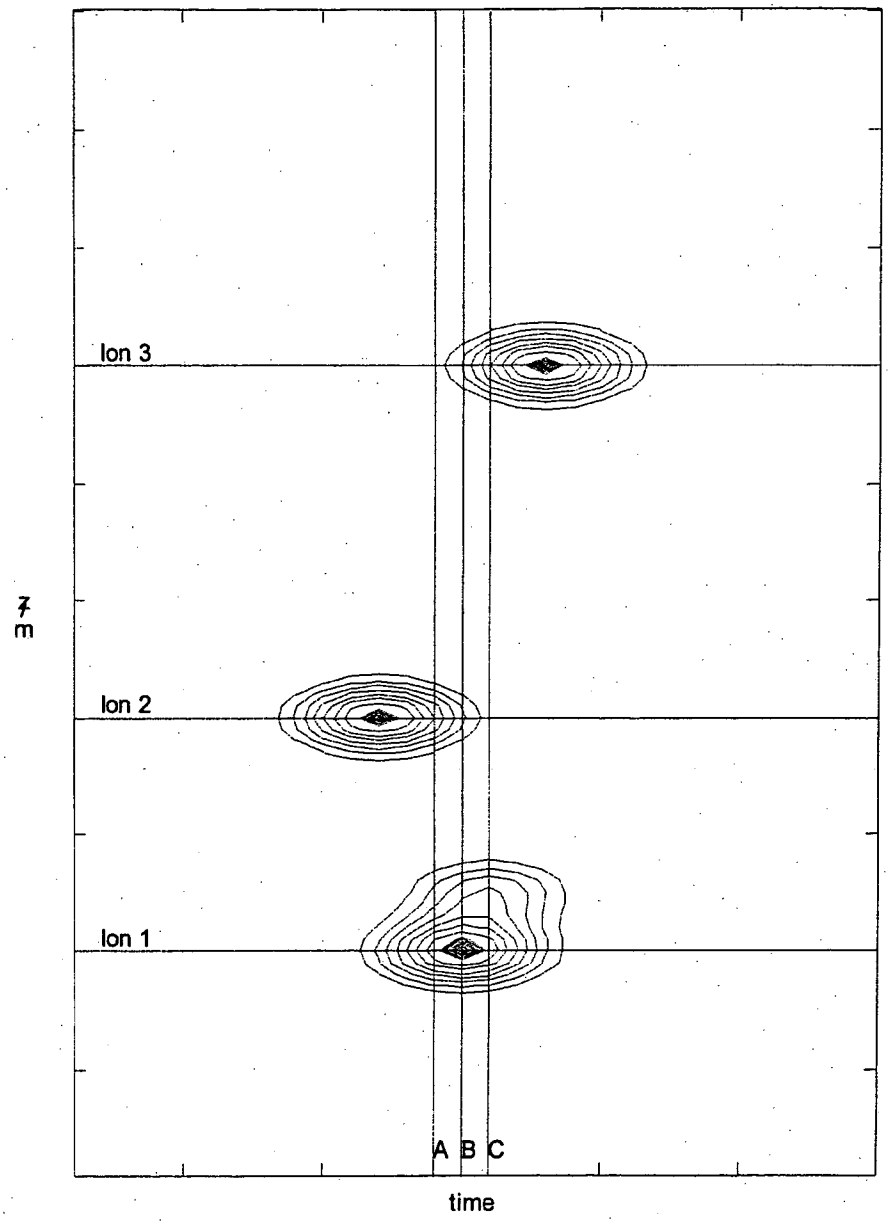


Figure 7. Example of coeluted ion in extracted spectra

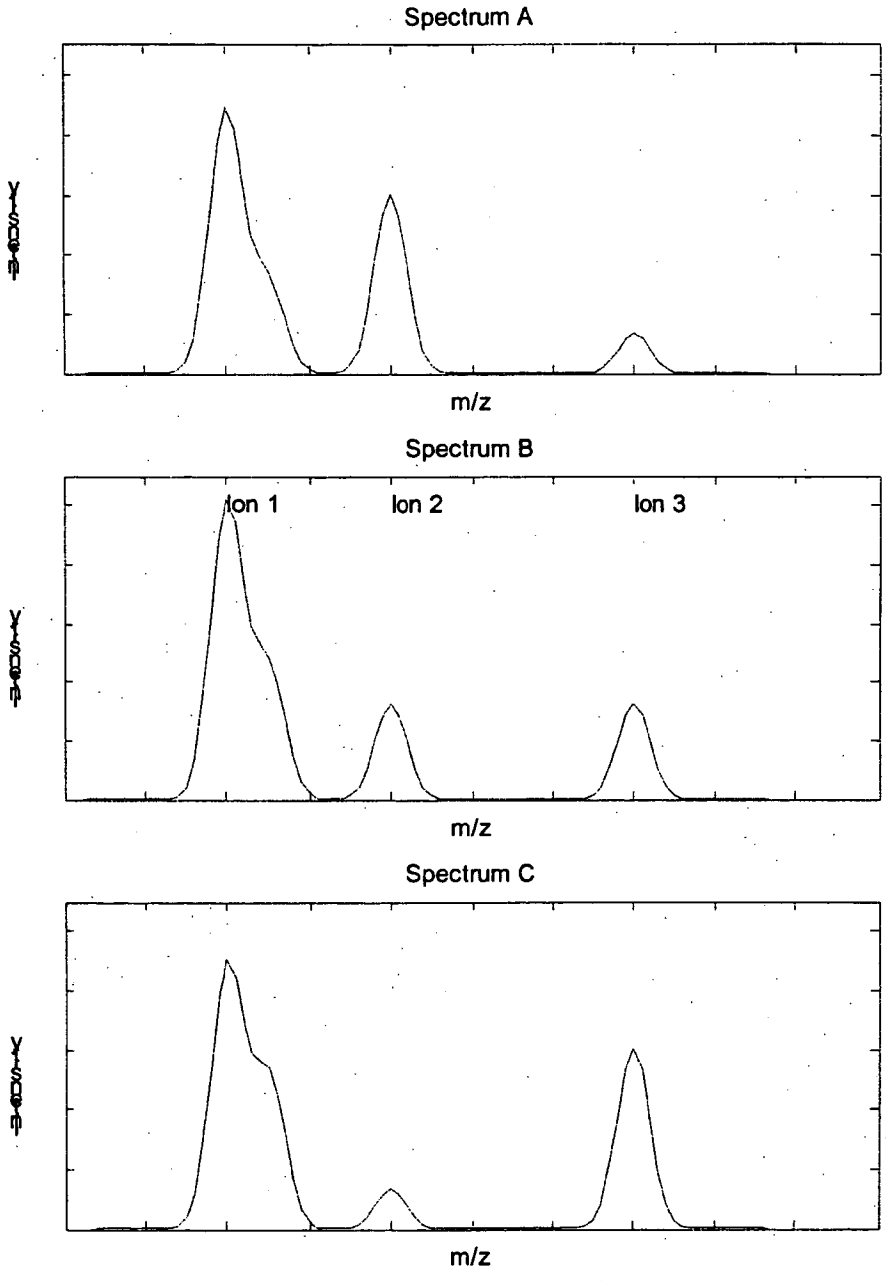


Figure 8. Example of ions in contour plot with nois

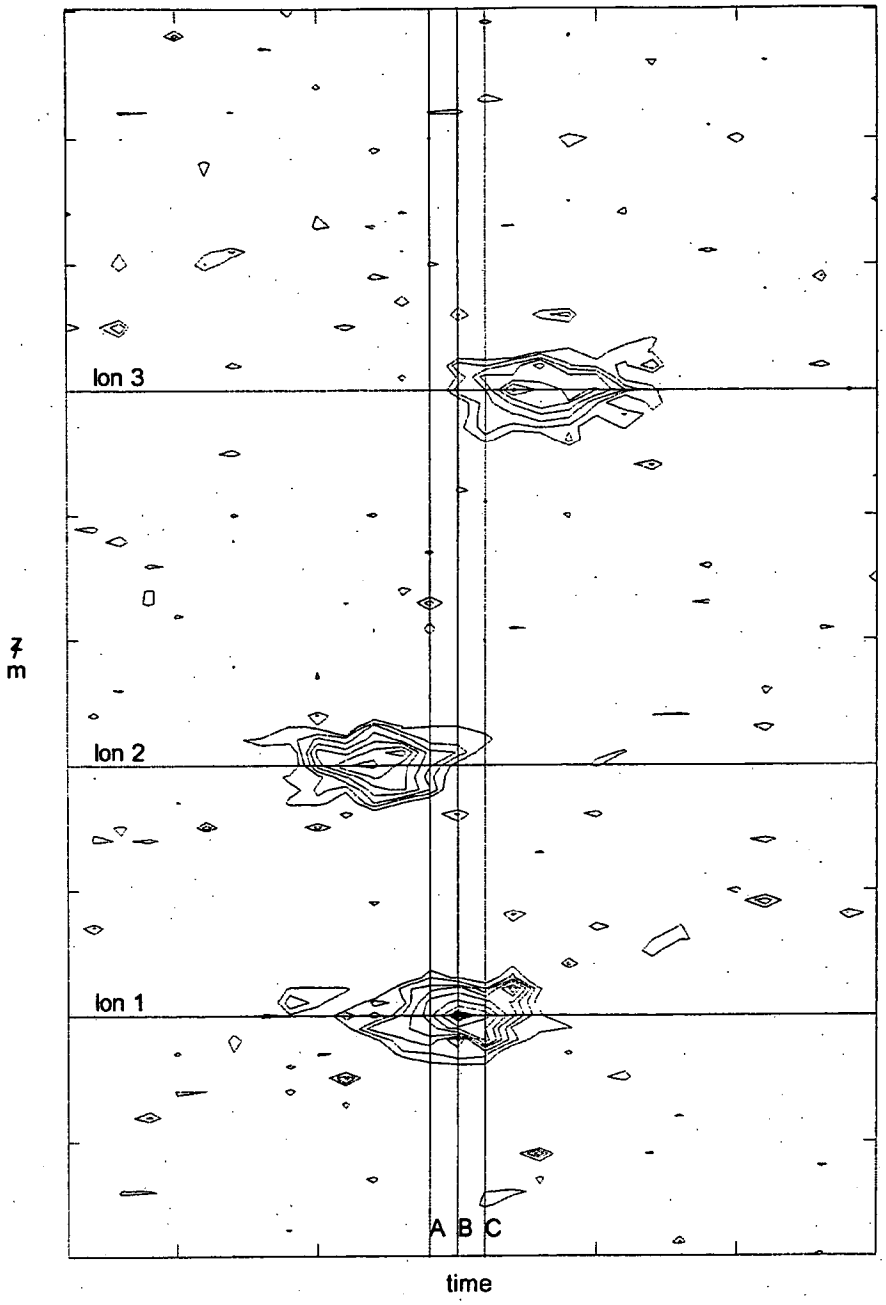


Figure 9. Example of ions in extracted spectra with noise

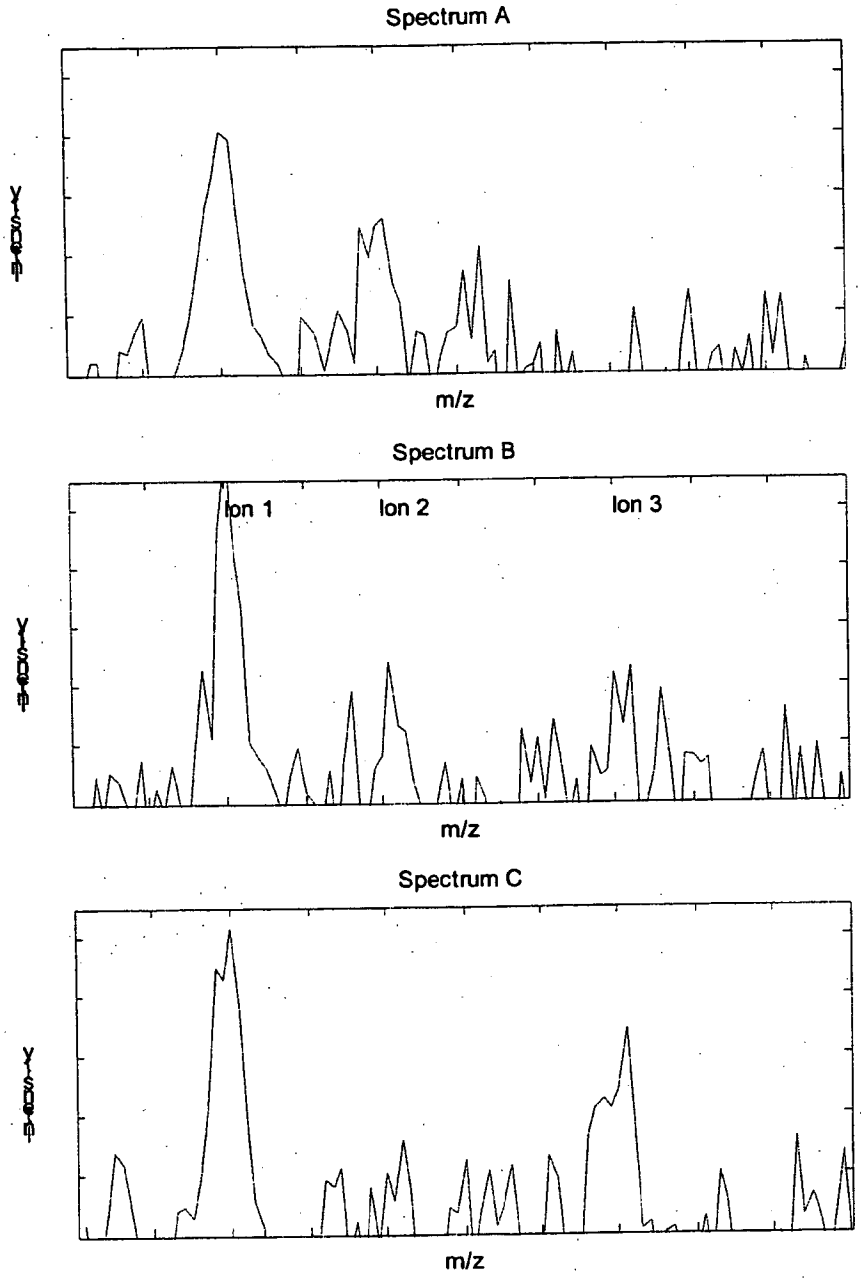


Figure 10. Example of ions in extracted chromatograms with noise

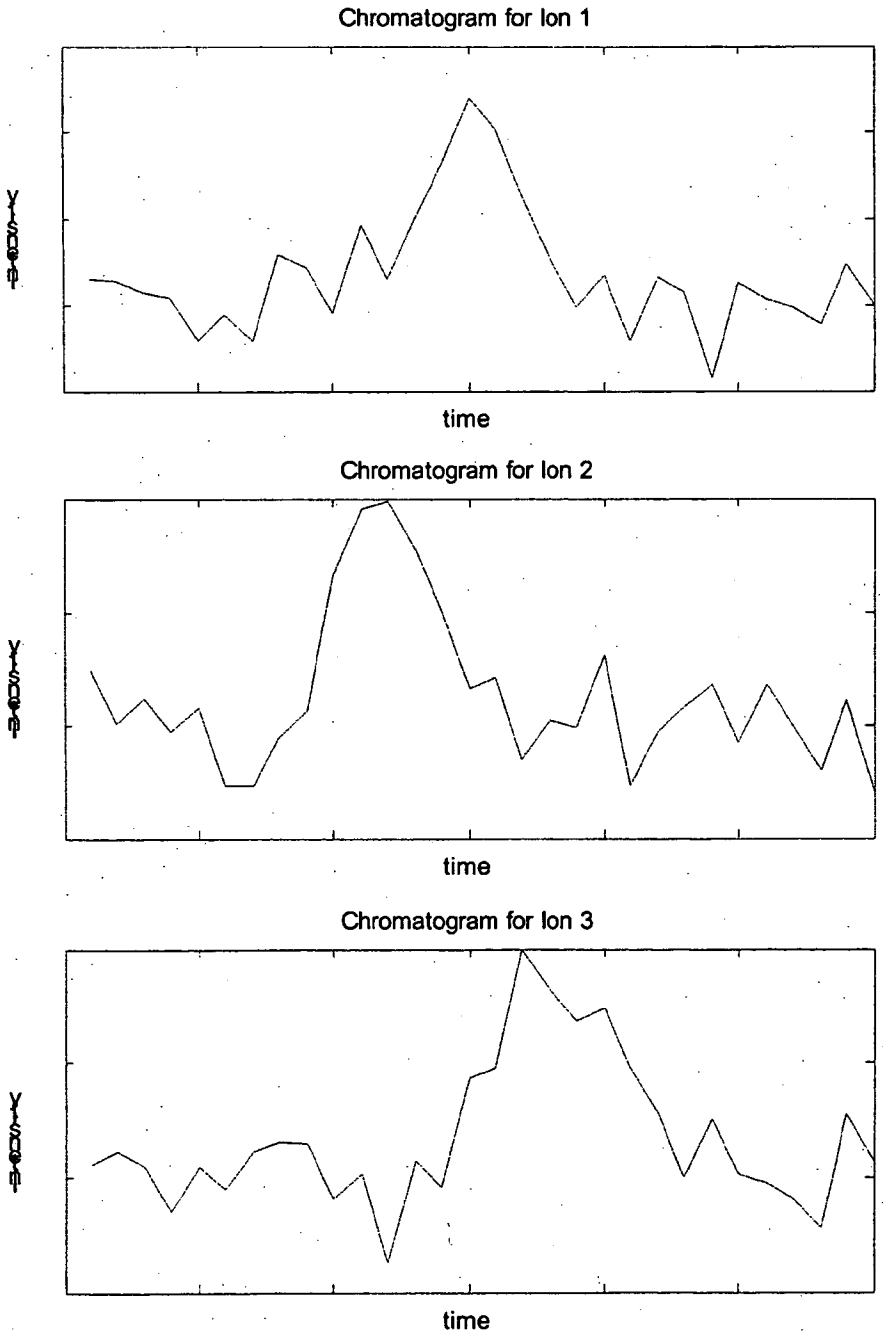


Figure 10.1 Ions after convolution

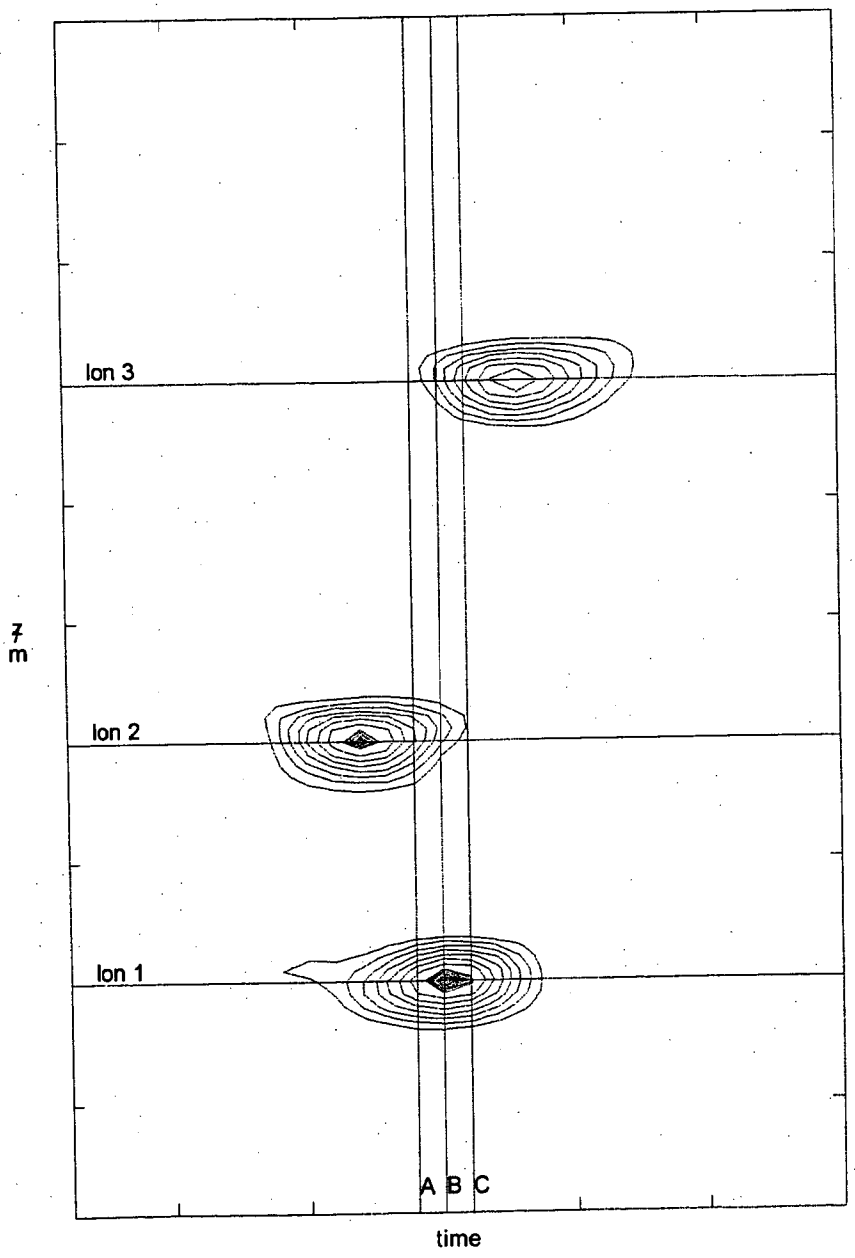


FIGURE 12. The ion detection and parameter estimation method

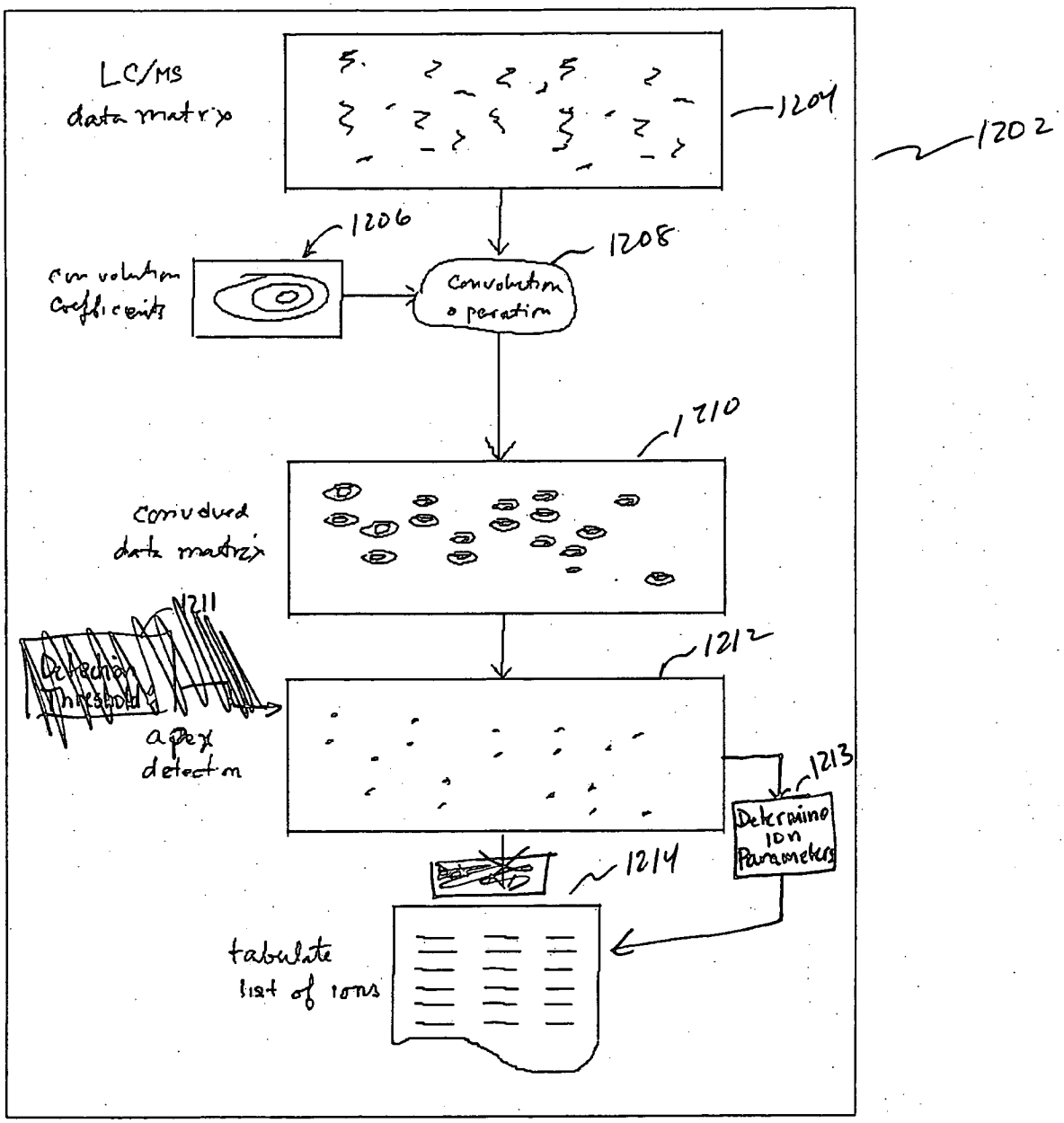


FIGURE 13 Threshold applied to ion list

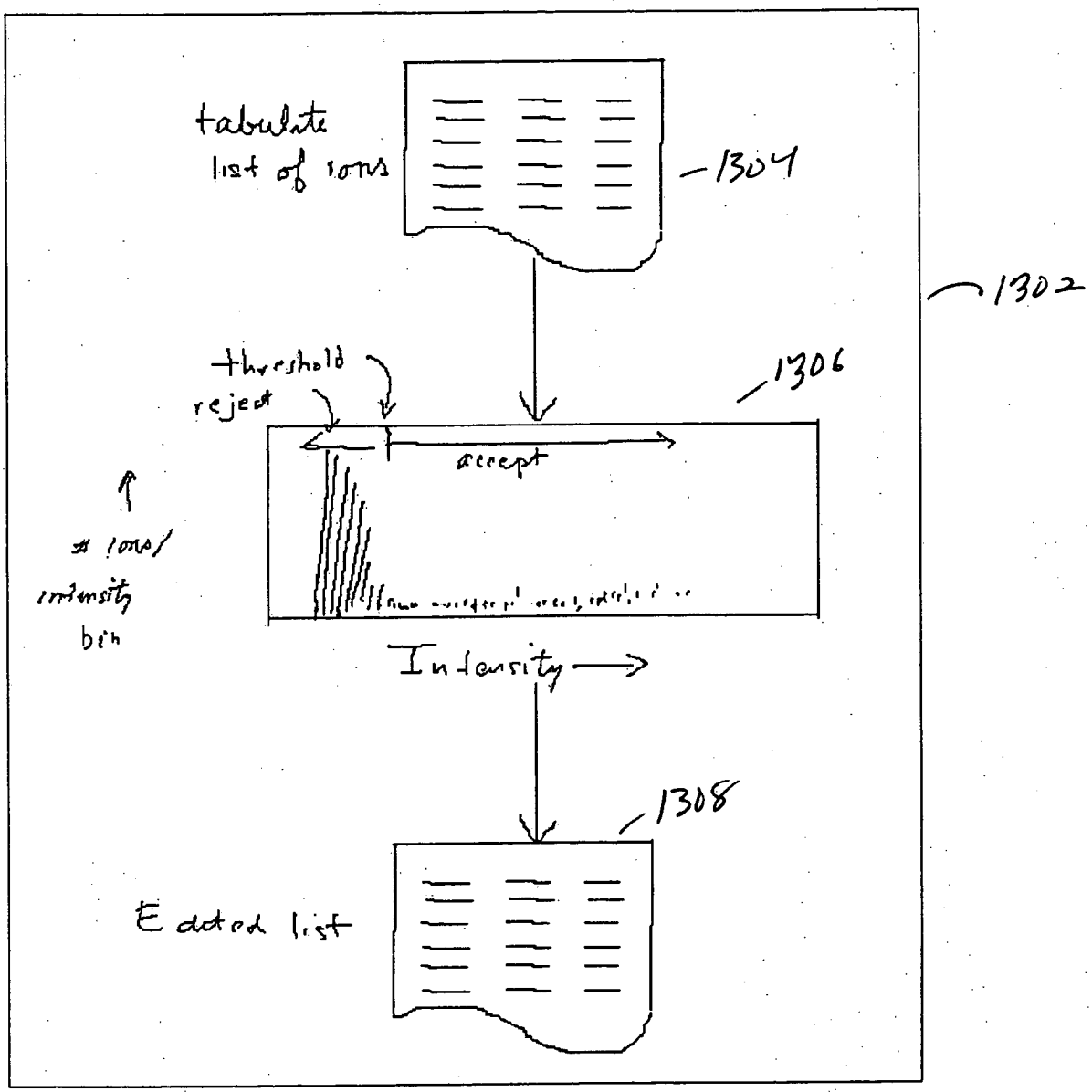
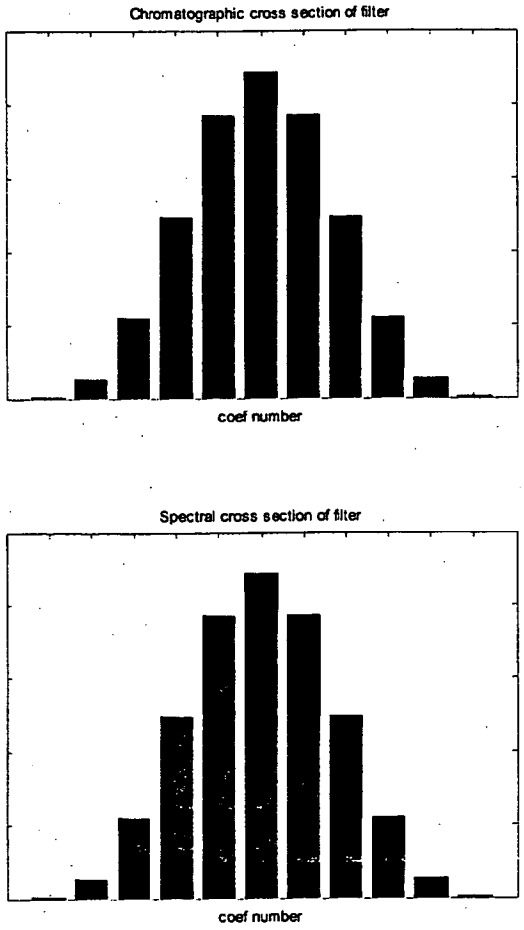


Figure 15 smoothing and second derivative filters for one-dimensional data

Smoothing



Second derivative

TBD

Figure 16. Example of two coeluting parent ~~ions~~ Molecules that each produce multiple ions

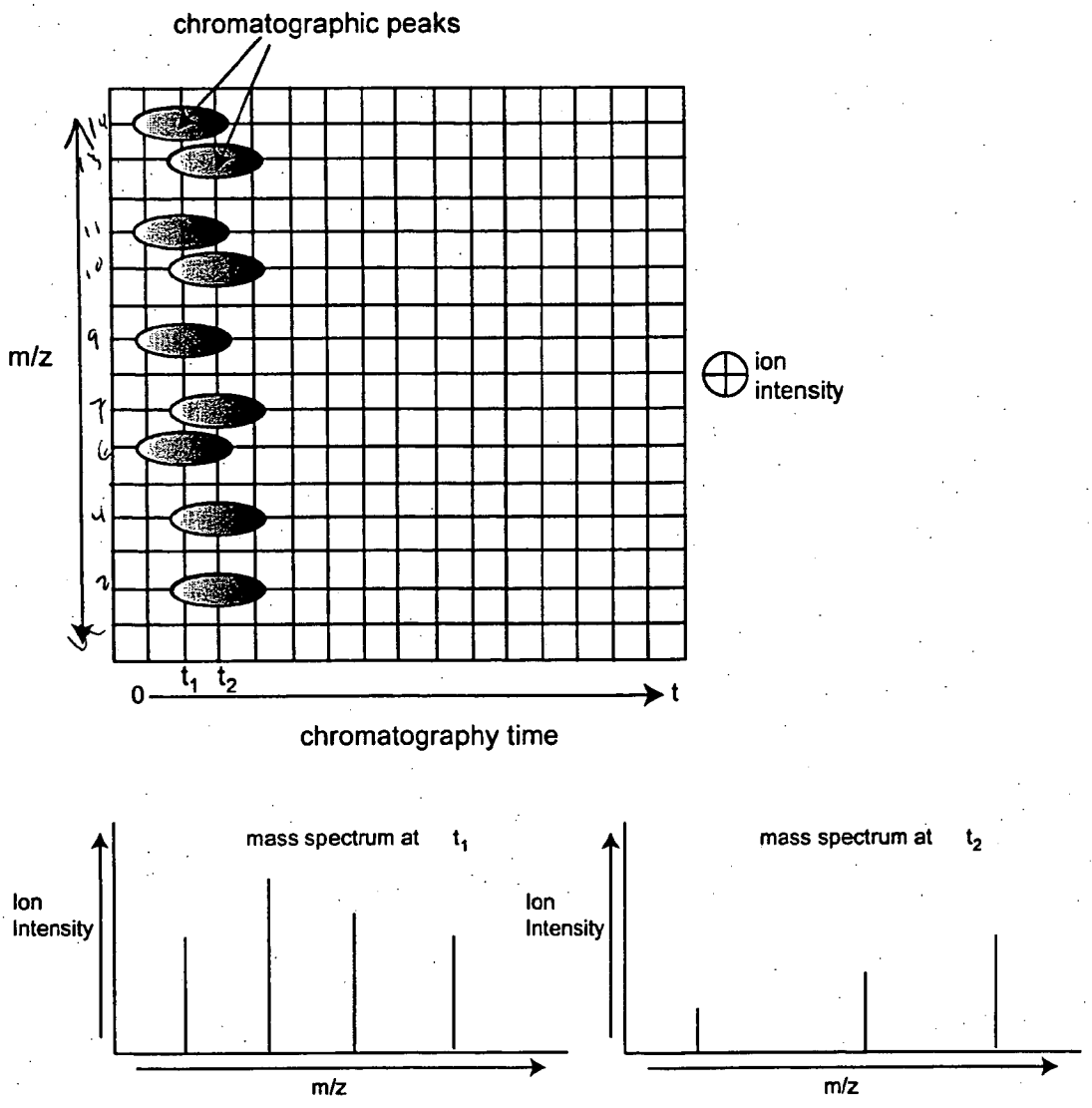


FIGURE 17 Fragmentation peaks that occur at retention times corresponding to precursor ions.